(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 02/42418 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US01/43607

(22) International Filing Date:
20 November 2001 (20.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

| | | |
|---|---|---|
| 60/252,123 | 20 November 2000 (20.11.2000) | US |
| 60/285,478 | 20 April 2001 (20.04.2001) | US |
| 60/306,727 | 20 July 2001 (20.07.2001) | US |
| 60/317,845 | 7 September 2001 (07.09.2001) | US |

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

| | |
|---|---|
| US | 60/252,123 (CON) |
| Filed on | 20 November 2000 (20.11.2000) |
| US | 60/285,478 (CON) |
| Filed on | 20 April 2001 (20.04.2001) |
| US | 60/306,727 (CON) |
| Filed on | 20 July 2001 (20.07.2001) |
| US | 60/317,845 (CON) |
| Filed on | 7 September 2001 (07.09.2001) |

(71) Applicant *(for all designated States except US)*: **CARGILL, INCORPORATED** [US/US]; 15407 McGinty Road West, Wayzata, MN 55391-2399 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **GOKARN, Ravi, R.** [—/US]; 3205 Harbor, Lane #4311, Plymouth, MN 55447 (US). **SELIFONOVA, Olga, V.** [RU/US]; 1405 Olive Lane N. #318, Plymouth, MN 55447 (US). **JESSEN, Holly** [US/US]; 6618 Brenden Court, Chanhassen, MN 55317 (US). **GORT, Steven, J.** [US/US]; 3207 Quarles Road, Brooklyn Park, MN 55429 (US). **SELMER, Thorsten** [DE/DE]; Cappeler Strasse 12, 35039 Marburg (DE). **BUCKEL, Wolfgang** [DE/DE]; Am Koeppel 8, 35043 Marburg (DE).

(74) Agent: **DEGRANDIS, Paula**; Cargill, Incorporated, P.O. Box 5624, Minneapolis, MN 55440-5624 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 02/42418 A2

(54) Title: 3-HYDROXYPROPIONIC ACID AND OTHER ORGANIC COMPOUNDS

(57) Abstract: The invention provides methods and materials related to producing 3-HP as well as other organic compounds. Specifically, the invention provides isolated nucleic acids, polypeptides, host cells, and methods and materials for producing 3-HP and other organic compounds.

# 3-HYDROXYPROPIONIC ACID AND OTHER ORGANIC COMPOUNDS

## FIELD OF THE INVENTION

The invention relates to enzymes and methods that can be used to produce organic acids and related products.

## CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority from the following U.S. Provisional Patent Applications, which are herein incorporated by reference: U.S. Provisional Patent Application Serial Number 60/252,123, filed November 20, 2000; U.S. Provisional Patent Application Serial Number 60/285,478, filed April 20, 2001; U.S. Provisional Patent Application Serial Number 60/306,727, filed July 20, 2001; and U.S. Provisional Patent Application Serial Number 60/317,845, filed September 7, 2001.

## BACKGROUND

Organic chemicals such as organic acids, esters, and polyols can be used to synthesize plastic materials and other products. To meet the increasing demand for organic chemicals, more efficient and cost effective production methods are being developed which utilize raw materials based on carbohydrates rather than hydrocarbons. For example, certain bacteria have been used to produce large quantities of lactic acid used in the production of polylactic acid.

3-hydroxypropionic acid (3-HP) is an organic acid. Although several chemical synthesis routes have been described to produce 3-HP, only one biocatalytic route has been heretofore previously disclosed (WO 01/16346 to Suthers, et al.). 3-HP has utility for specialty synthesis and can be converted to commercially important intermediates by known art in the chemical industry, e.g., acrylic acid by dehydration, malonic acid by

oxidation, esters by esterification reactions with alcohols, and reduction to 1,3 propanediol.

## SUMMARY

The invention relates to methods and materials involved in producing 3-hydroxypropionic acid and other organic compounds (e.g., 1,3-propanediol, acrylic acid, polymerized acrylate, esters of acrylate, polymerized 3-HP, esters of 3-HP, and malonic acid and its esters). Specifically, the invention provides nucleic acid molecules, polypeptides, host cells, and methods that can be used to produce 3-HP and other organic compounds such as 1,3-propanediol, acrylic acid, polymerized acrylate, esters of acrylate, polymerized 3-HP, esters of 3-HP, and malonic acid and its esters. 3-HP has potential to be both biologically and commercially important. For example, the nutritional industry can use 3-HP as a food, feed additive or preservative, while the derivatives mentioned above can be produced from 3-HP. The nucleic acid molecules described herein can be used to engineer host cells with the ability to produce 3-HP as well as other organic compounds such as 1,3-propanediol, acrylic acid, polymerized acrylate, esters of acrylate, polymerized 3-HP, and esters of 3-HP. The polypeptides described herein can be used in cell-free systems to make 3-HP as well as other organic compounds such as 1,3-propanediol, acrylic acid, polymerized acrylate, esters of acrylate, polymerized 3-HP, and esters of 3-HP. The host cells described herein can be used in culture systems to produce large quantities of 3-HP as well as other organic compounds such as 1,3-propanediol, acrylic acid, polymerized acrylate, esters of acrylate, polymerized 3-HP, and esters of 3-HP.

One aspect of the invention provides cells that have lactyl-CoA dehydratase activity and 3-hydroxypropionyl-CoA dehydratase activity, and methods of making products such as those described herein by culturing at least one of the cells that have lactyl-CoA dehydratase activity and 3-hydroxypropionyl-CoA dehydratase activity. In some embodiments, the cell can also contain an exogenous nucleic acid molecule that encodes one or more of the following polypeptides: a polypeptide having E1 activator activity; an E2 α polypeptide that is a subunit of an enzyme having lactyl-CoA dehydratase activity; an E2 β polypeptide that is a subunit of an enzyme having lactyl-

CoA dehydratase activity; and a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity. Additionally, the cell can have CoA transferase activity, CoA synthetase activity, poly hydroxyacid synthase activity, 3-hydroxypropionyl-CoA hydrolase activity, 3-hydroxyisobutryl-CoA hydrolase activity, and/or lipase activity. Moreover, the cell can contain at least one exogenous nucleic acid molecule that expresses one or more polypeptides that have CoA transferase activity, 3-hydroxypropionyl-CoA hydrolase activity, 3-hydroxyisobutryl-CoA hydrolase activity, CoA synthetase activity, poly hydroxyacid synthase activity, and/or lipase activity.

In another embodiment of the invention, the cell that has lactyl-CoA dehydratase activity and 3-hydroxypropionyl-CoA dehydratase activity produces a product, for example, 3-HP, polymerized 3-HP, and/or an ester of 3-HP, such as methyl hydroxypropionate, ethyl hydroxypropionate, propyl hydroxypropionate, and/or butyl hydroxypropionate. Accordingly, the invention also provides methods of producing one or more of these products. These methods involve culturing the cell that has lactyl-CoA dehydratase activity and 3-hydroxypropionyl-CoA dehydratase activity under conditions that allow the product to be produced. These cells also can have CoA synthetase activity and/or poly hydroxyacid synthase activity.

Another aspect of the invention provides cells that have CoA synthetase activity, lactyl-CoA dehydratase activity, and poly hydroxyacid synthase activity. In some embodiments, these cells also can contain an exogenous nucleic acid molecule that encodes one or more of the following polypeptides: a polypeptide having E1 activator activity; an E2 α polypeptide that is a subunit of an enzyme having lactyl-CoA dehydratase activity; an E2 β polypeptide that is a subunit of an enzyme having lactyl-CoA dehydratase activity; a polypeptide having CoA synthetase activity; and a polypeptide having poly hydroxyacid synthase activity.

In another embodiment of the invention, the cell that has CoA synthetase activity, lactyl-CoA dehydratase activity, and poly hydroxyacid synthase activity can produce a product, for example, polymerized acrylate.

Another aspect of the invention provides a cell comprising CoA transferase activity, lactyl-CoA dehydratase activity, and lipase activity. In some embodiments, the cell also can contain an exogenous nucleic acid molecule that encodes one or more of the

following polypeptides: a polypeptide having CoA transferase activity; a polypeptide having E1 activator activity; an E2 α polypeptide that is a subunit of an enzyme having lactyl-CoA dehydratase activity; an E2 β polypeptide that is a subunit of an enzyme having lactyl-CoA dehydratase activity; and a polypeptide having lipase activity. This cell can be used, among other things, to produce products such as esters of acrylate (e.g., methyl acrylate, ethyl acrylate, propyl acrylate, and butyl acrylate).

In some embodiments, 1,3 propanediol can be created from either 3-HP-CoA or 3-HP via the use of polypeptides having enzymatic activity. These polypeptides can be used either *in vitro* or *in vivo*. When converting 3-HP-CoA to 1,3 propanediol, polypeptides having oxidoreductase activity or reductase activity (e.g., enzymes from the 1.1.1.- class of enzymes) can be used. Alternatively, when creating 1,3 propanediol from 3-HP, a combination of (1) a polypeptide having aldyhyde dehydrogenase activity (e.g., an enzyme from the 1.1.1.34 class) and (2) a polypeptide having alcohol dehydrogenase activity (e.g., an enzyme from the 1.1.1.32 class) can be used.

In some embodiments of the invention, products are produced *in vitro* (outside of a cell). In other embodiments of the invention, products are produced using a combination of *in vitro* and *in vivo* (within a cell) methods. In yet other embodiments of the invention, products are produced *in vivo*. For methods involving *in vivo* steps, the cells can be isolated cultured cells or whole organisms such as transgenic plants, non-human mammals, or single-celled organisms such as yeast and bacteria (e.g., *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells). Hereinafter such cells are referred to as production cells. Products produced by these production cells can be organic products such as 3-HP and/or the nucleic acid molecules and polypeptides described herein.

Another aspect of the invention provides polypeptides having an amino acid sequence that (1) is set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161, (2) is at least 10 contiguous amino acid residues of a sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161, (3) has at least 65 percent sequence identity with at least 10 contiguous amino acid residues of a sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161, (4) is a sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161 having conservative amino acid substitutions,

or (5) has at least 65 percent sequence identity with a sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. Accordingly, the invention also provides nucleic acid sequences that encode any of the polypeptides described herein as well as specific binding agents that bind to any of the polypeptides described herein. Likewise, the invention provides transformed cells that contain any of the nucleic acid sequences that encode any of the polypeptides described herein. These cells can be used to produce nucleic acid molecules, polypeptides, and organic compounds. The polypeptides can be used to catalyze the formation of organic compounds or can be used as antigens to create specific binding agents.

In yet another embodiment, the invention provides isolated nucleic acid molecules that contain at least one of the following nucleic acid sequences: (1) a nucleic acid sequence as set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163; (2) a nucleic acid sequence having at least 10 consecutive nucleotides from a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163; (3) a nucleic acid sequences that hybridize under hybridization conditions (e.g., moderately or highly stringent hybridization conditions) to a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163; (4) a nucleic acid sequence having 65 percent sequence identity with at least 10 consecutive nucleotides from a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163; and (5) a nucleic acid sequence having at least 65 percent sequence identity with a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163. Accordingly, the invention also provides a production cell that contains at least one exogenous nucleic acid having any the nucleic acid sequences provided above. The production cell can be used to express polypeptides that have an enzymatic activity such as CoA transferase activity, lactyl-CoA dehydratase activity, CoA synthase activity, dehydratase activity, dehydrogenase activity, malonyl CoA reductase activity, β-alanine ammonia lyase activity, and/or 3-hydroxypropionyl-CoA dehydratase activity. Accordingly, the invention also provides methods of producing polypeptides encoded by the nucleic acid sequences described above.

The invention also provides several methods such as methods for making 3-HP from lactate, phosphoenolpyruvate (PEP), or pyruvate. In some embodiments, methods

for making 3-HP from lactate, PEP, or pyruvate involve culturing a cell containing at least one exogenous nucleic acid under conditions that allow the cell to produce 3-HP. These methods can be practiced using the various types of production cells described herein. In some embodiments, the production cells can have one or more of the following activities: CoA transferase activity, 3-hydroxypropionyl-CoA hydrolase activity, 3-hydroxyisobutryl-CoA hydrolase activity, dehydratase activity, and/or malonyl CoA reductase activity.

In other embodiments, the methods involve making 3-HP wherein lactate is contacted with a first polypeptide having CoA transferase activity or CoA synthetase activity such that lactyl-CoA is formed, then contacting lactyl-CoA with a second polypeptide having lactyl-CoA dehydratase activity to form acrylyl-CoA, then contacting acrylyl-CoA with a third polypeptide having 3-hydroxypropionyl-CoA dehydratase activity to form 3-hydroxypropionic acid-CoA, and then contacting 3-hydroxypropionic acid-CoA with the first polypeptide to form 3-HP or with a fourth polypeptide having 3-hydroxypropionyl-CoA hydrolase activity or 3-hydroxyisobutryl-CoA hydrolase activity to form 3-HP.

Another aspect of the invention provides methods for making polymerized 3-HP. These methods involve making 3-hydroxypropionic acid-CoA as described above, and then contacting the 3-hydroxypropionic acid-CoA with a polypeptide having poly hydroxyacid synthase activity to form polymerized 3-HP.

In yet another embodiment of the invention, methods for making an ester of 3-HP are provided. These methods involve making 3-HP as described above, and then additionally contacting 3-HP with a fifth polypeptide having lipase activity to form an ester.

The invention also provides methods for making polymerized acrylate. These methods involve culturing a cell that has both CoA synthetase activity, lactyl-CoA dehydratase activity, and poly hydroxyacid synthase activity such that polymerized acrylate is made. Accordingly, the invention also provides methods of making polymerized acrylate wherein lactate is contacted with a first polypeptide having CoA synthetase activity to form lactyl-CoA, then contacting lactyl-CoA with a second polypeptide having lactyl-CoA dehydratase activity to form acrylyl-CoA, and then

contacting acrylyl-CoA with a third polypeptide having poly hydroxyacid synthase activity to form polymerized acrylate.

The invention also provides methods of making an ester of acrylate. These methods involve culturing a cell that has CoA transferase activity, lipase activity, and lactyl-CoA dehydratase activity under conditions that allow the cell to produce an ester.

In another embodiment, the invention provides methods for making an ester of acrylate, wherein acrylyl-CoA is formed as described above, and then acrylyl-CoA is contacted with a polypeptide having CoA transferase activity to form acrylate, and acrylate is contacted with a polypeptide having lipase activity to form the ester.

The invention also provides methods for making 3-HP. These methods involve culturing a cell containing at least one exogenous nucleic acid that encodes at least one polypeptide such that 3-HP is produced from acetyl-CoA or malonyl-CoA.

Alternative embodiments provide methods of making 3-HP, wherein acetyl-CoA is contacted with a first polypeptide having acetyl-CoA carboxylase activity to form malonyl-CoA, and malonyl-CoA is contacted with a second polypeptide having malonyl-CoA reductase activity to form 3-HP.

In other embodiments, malonyl-CoA can be contacted with a polypeptide having malonyl-CoA reductase activity so that 3-HP can be made.

In another embodiment, the invention provides a method for making 3-HP that uses a β-alanine intermediate. This method can be performed by contacting β-alanine CoA with a first polypeptide having β-alanyl-CoA ammonia lyase activity (such as a polypeptide having the amino acid sequence set forth in SEQ ID NO: 160 or 161) to form acrylyl-CoA, contacting acrylyl-CoA with a second polypeptide having 3-HP-CoA dehydratase activity to form 3-HP-CoA, and contacting 3-HP-CoA with a third polypeptide having glutamate dehydrogenase activity to make 3-HP.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In

case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

## DESCRIPTION OF DRAWINGS

Figure 1 is a diagram of a pathway for making 3-HP.

Figure 2 is a diagram of a pathway for making polymerized 3-HP.

Figure 3 is a diagram of a pathway for making esters of 3-HP.

Figure 4 is a diagram of a pathway for making polymerized acrylic acid.

Figure 5 is a diagram of a pathway for making esters of acrylate.

Figure 6 is a listing of a nucleic acid sequence that encodes a polypeptide having CoA transferase activity (SEQ ID NO:1).

Figure 7 is a listing of an amino acid sequence of a polypeptide having CoA transferase activity (SEQ ID NO:2).

Figure 8 is an alignment of the nucleic acid sequences set forth in SEQ ID NOs:1, 3, 4, and 5.

Figure 9 is an alignment of the amino acid sequences set forth in SEQ ID NOs:2, 6, 7, and 8.

Figure 10 is a listing of a nucleic acid sequence that encodes a polypeptide having E1 activator activity (SEQ ID NO:9).

Figure 11 is a listing of an amino acid sequence of a polypeptide having E1 activator activity (SEQ ID NO:10).

Figure 12 is an alignment of the nucleic acid sequences set forth in SEQ ID NOs:9, 11, 12, and 13.

Figure 13 is an alignment of the amino acid sequences set forth in SEQ ID NOs:10, 14, 15, and 16.

Figure 14 is a listing of a nucleic acid sequence that encodes an E2 α subunit of an enzyme having lactyl-CoA dehydratase activity (SEQ ID NO:17).

Figure 15 is a listing of an amino acid sequence of an E2 α subunit of an enzyme having lactyl-CoA dehydratase activity (SEQ ID NO:18).

Figure 16 is an alignment of the nucleic acid sequences set forth in SEQ ID NOs:17, 19, 20, and 21.

Figure 17 is an alignment of the amino acid sequences set forth in SEQ ID NOs:18, 22, 23, and 24.

Figure 18 is a listing of a nucleic acid sequence that encodes an E2 β subunit of an enzyme having lactyl-CoA dehydratase activity (SEQ ID NO:25). The "G" at position 443 can be an "A"; and the "A" at position 571 can be a "G".

Figure 19 is a listing of an amino acid sequence of an E2 β subunit of an enzyme having lactyl-CoA dehydratase activity (SEQ ID NO:26).

Figure 20 is an alignment of the nucleic acid sequences set forth in SEQ ID NOs:25, 27, 28, and 29.

Figure 21 is an alignment of the amino acid sequences set forth in SEQ ID NOs:26, 30, 31, and 32.

Figure 22 is a listing of a nucleic acid sequence of genomic DNA from *Megasphaera elsdenii* (SEQ ID NO:33).

Figure 23 is a listing of a nucleic acid sequence that encodes a polypeptide from *Megasphaera elsdenii* (SEQ ID NO:34).

Figure 24 is a listing of an amino acid sequence of a polypeptide from *Megasphaera elsdenii* (SEQ ID NO:35).

Figure 25 is a listing of a nucleic acid sequence that encodes a polypeptide having enzymatic activity (SEQ ID NO:36).

Figure 26 is a listing of an amino acid sequence of a polypeptide having enzymatic activity (SEQ ID NO:37).

Figure 27 is a listing of a nucleic acid sequence that contains non-coding as well as coding sequence of a polypeptide having CoA synthase, dehydratase, and dehydrogenase activity (SEQ ID NO:38). The start site for the coding sequence is at position 480, a ribosome binding site is at position 466-473, and the stop codon is at position 5946.

Figure 28 is a listing of an amino acid sequence from a polypeptide having CoA synthase, dehydratase, and dehydrogenase activity (SEQ ID NO:39).

Figure 29 is a listing of a nucleic acid sequence that encodes a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity (SEQ ID NO:40).

Figure 30 is a listing of an amino acid sequence of a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity (SEQ ID NO:41).

Figure 31 is a listing of a nucleic acid sequence that contains non-coding as well as coding sequence of a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity (SEQ ID NO:42).

Figure 32 is an alignment of the nucleic acid sequences set forth in SEQ ID NOs:40, 43, 44, and 45.

Figure 33 is an alignment of the amino acid sequences set forth in SEQ ID NOs:41, 46, 47, and 48.

Figure 34 is a diagram of the construction of a synthetic operon (pTDH) that encodes for polypeptides having CoA transferase activity, lactyl-CoA dehydratase activity (E1, E2 α, and E2 β), and 3-hydroxypropionyl-CoA dehydratase activity (3-HP-CoA dehydratase).

Figure 35A and B is a diagram of the construction of a synthetic operon (pHTD) that encodes for polypeptides having CoA transferase activity, lactyl-CoA dehydratase activity (E1, E2 α, and E2 β), and 3-hydroxypropionyl-CoA dehydratase activity (3-HP-CoA dehydratase).

Figure 36A and B is a diagram of the construction of a synthetic operon (pEIITHrEI) that encodes for polypeptides having CoA transferase activity, lactyl-CoA dehydratase activity (E1, E2 α, and E2 β), and 3-hydroxypropionyl-CoA dehydratase activity (3-HP-CoA dehydratase).

Figure 37A and B is a diagram of the construction of a synthetic operon (pEIITHEI) that encodes for polypeptides having CoA transferase activity, lactyl-CoA dehydratase activity (E1, E2 α, and E2 β), and 3-hydroxypropionyl-CoA dehydratase activity (3-HP-CoA dehydratase).

Figure 38A and B is a diagram of the construction of two plasmids, pEIITH and pPROEI. The pEIITH plasmid encodes polypeptides having CoA transferase activity, lactyl-CoA dehydratase activity (E2 α and E2 β), and 3-hydroxypropionyl-CoA

dehydratase activity (3-HP-CoA dehydratase), and the pPROEI plasmid encodes a polypeptide having E1 activator activity.

Figure 39 is a listing of a nucleic acid sequence that encodes a polypeptide having CoA synthase, dehydratase, and dehydrogenase activity (SEQ ID NO:129).

Figure 40 is an alignment of the amino acid sequences set forth in SEQ ID NOs:39, 130, and 131. The uppercase amino acid residues represent positions where that amino acid residue is present in two or more sequences.

Figure 41 is an alignment of the amino acid sequences set forth in SEQ ID NOs:39, 132, and 133. The uppercase amino acid residues represent positions where that amino acid residue is present in two or more sequences.

Figure 42 is an alignment of the amino acid sequences set forth in SEQ ID NOs: 39, 134, and 135. The uppercase amino acid residues represent positions where that amino acid residue is present in two or more sequences.

Figure 43 is a diagram of several pathways for making organic compounds using the multifunctional OS17 enzyme.

Figure 44 is a diagram of a pathway for making 3-HP via acetyl-CoA and malonyl-CoA.

Figure 45 is a diagram of pMSD8, pET30a/acc1, pFN476, and PET286 constructs.

Figure 46 contains a total ion chromatogram and five mass spectrums of Coenzyme A thioesters. Panel A is total ion chromatogram illustrating the separation of Coenzyme A and four CoA-organic thioesters: 1=Coenzyme A, 2=lactyl-CoA, 3=acetyl-CoA, 4=acrylyl-CoA, 5=propionyl-CoA. Panel B is a mass spectrum of Coenzyme A. Panel C is a mass spectrum of lactyl-CoA. Panel D is a mass spectrum of acetyl-CoA. Panel E is a mass spectrum of acrylyl-CoA. Panel F is a mass spectrum of propionyl-CoA.

Figure 47 contains ion chromatograms and mass spectrums. Panel A is a total ion chromatogram of a mixture of lactyl-CoA and 3-HP-CoA. The Panel A insert is the mass spectrum recorded under peak 1. Panel B is a total ion chromatogram of lactyl-CoA. The Panel B insert is the mass spectrum recorded under peak 2. In each panel, peak 1 is 3-HP-CoA, and peak 2 is lactyl-CoA. The peak labeled with an asterisk was confirmed not to be a CoA ester.

Figure 48 contains ion chromatograms and mass spectrums. Panel A is a total ion chromatogram of CoA esters derived from a broth produced by *E. coli* transfected with pEIITHrEI. The Panel A insert is the mass spectrum recorded under peak 1. Panel B is a total ion chromatogram of CoA esters derived from a broth produced by control *E. coli* not transfected with pEIITHrEI. The Panel B insert is the mass spectrum recorded under peak 2. In each panel, peak 1 is 3-HP-CoA, and peak 2 is lactyl-CoA. The peaks labeled with an asterisk were confirmed not to be a CoA ester.

Figure 49 is a listing of a nucleic acid sequence that encodes a polypeptide having malonyl-CoA reductase activity (SEQ ID NO: 140).

Figure 50 is a listing of an amino acid sequence of a polypeptide having malonyl-CoA reductase activity (SEQ ID NO:141).

Figure 51 is a listing of a nucleic acid sequence that encodes a portion of a polypeptide having malonyl-CoA reductase activity (SEQ ID NO:142).

Figure 52 is an alignment of the amino acid sequences set forth in SEQ ID NOs: 141, 143, 144, 145, 146, and 147.

Figure 53 is an alignment of the nucleic acid sequences set forth in SEQ ID NOs: 140, 148, 149, 150, 151, and 152.

Figure 54 is a diagram of a pathway for making 3-HP via a β-alanine intermediate.

Figure 55 is a diagram of a pathway for making 3-HP via a β-alanine intermediate.

Figure 56 is a listing of an amino acid sequence of a polypeptide having β-alanyl-CoA ammonia lyase activity (SEQ ID NO:160).

Figure 57 is a listing of an amino acid sequence of a polypeptide having β-alanyl-CoA ammonia lyase activity (SEQ ID NO:161).

Figure 58 is a listing of a nucleic acid sequence that encodes a polypeptide having β-alanyl-CoA ammonia lyase activity (SEQ ID NO:162).

Figure 59 is a listing of a nucleic acid sequence that can encode a polypeptide having β-alanyl-CoA ammonia lyase activity (SEQ ID NO:163).

## DETAILED DESCRIPTION

### I. Terms

**Nucleic acid:** The term "nucleic acid" as used herein encompasses both RNA and DNA including, without limitation, cDNA, genomic DNA, and synthetic (e.g., chemically synthesized) DNA. The nucleic acid can be double-stranded or single-stranded. Where single-stranded, the nucleic acid can be the sense strand or the antisense strand. In addition, nucleic acid can be circular or linear.

**Isolated:** The term "isolated" as used herein with reference to nucleic acid refers to a naturally-occurring nucleic acid that is not immediately contiguous with both of the sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally-occurring genome of the organism from which it is derived. For example, an isolated nucleic acid can be, without limitation, a recombinant DNA molecule of any length, provided one of the nucleic acid sequences normally found immediately flanking that recombinant DNA molecule in a naturally-occurring genome is removed or absent. Thus, an isolated nucleic acid includes, without limitation, a recombinant DNA that exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences as well as recombinant DNA that is incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., a retrovirus, adenovirus, or herpes virus), or into the genomic DNA of a prokaryote or eukaryote. In addition, an isolated nucleic acid can include a recombinant DNA molecule that is part of a hybrid or fusion nucleic acid sequence.

The term "isolated" as used herein with reference to nucleic acid also includes any non-naturally-occurring nucleic acid since non-naturally-occurring nucleic acid sequences are not found in nature and do not have immediately contiguous sequences in a naturally-occurring genome. For example, non-naturally-occurring nucleic acid such as an engineered nucleic acid is considered to be isolated nucleic acid. Engineered nucleic acid can be made using common molecular cloning or chemical nucleic acid synthesis techniques. Isolated non-naturally-occurring nucleic acid can be independent of other sequences, or incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., a retrovirus, adenovirus, or herpes virus), or the genomic DNA of a prokaryote or

eukaryote. In addition, a non-naturally-occurring nucleic acid can include a nucleic acid molecule that is part of a hybrid or fusion nucleic acid sequence.

It will be apparent to those of skill in the art that a nucleic acid existing among hundreds to millions of other nucleic acid molecules within, for example, cDNA or genomic libraries, or gel slices containing a genomic DNA restriction digest is not to be considered an isolated nucleic acid.

**Exogenous:** The term "exogenous" as used herein with reference to nucleic acid and a particular cell refers to any nucleic acid that does not originate from that particular cell as found in nature. Thus, non-naturally-occurring nucleic acid is considered to be exogenous to a cell once introduced into the cell. Nucleic acid that is naturally-occurring also can be exogenous to a particular cell. For example, an entire chromosome isolated from a cell of person X is an exogenous nucleic acid with respect to a cell of person Y once that chromosome is introduced into Y's cell.

**Hybridization:** The term "hybridization" as used herein refers to a method of testing for complementarity in the nucleotide sequence of two nucleic acid molecules, based on the ability of complementary single-stranded DNA and/or RNA to form a duplex molecule. Nucleic acid hybridization techniques can be used to obtain an isolated nucleic acid within the scope of the invention. Briefly, any nucleic acid having some homology to a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163 can be used as a probe to identify a similar nucleic acid by hybridization under conditions of moderate to high stringency. Once identified, the nucleic acid then can be purified, sequenced, and analyzed to determine whether it is within the scope of the invention as described herein.

Hybridization can be done by Southern or Northern analysis to identify a DNA or RNA sequence, respectively, that hybridizes to a probe. The probe can be labeled with a biotin, digoxygenin, an enzyme, or a radioisotope such as $^{32}P$. The DNA or RNA to be analyzed can be electrophoretically separated on an agarose or polyacrylamide gel, transferred to nitrocellulose, nylon, or other suitable membrane, and hybridized with the probe using standard techniques well known in the art such as those described in sections 7.39-7.52 of Sambrook et al., (1989) Molecular Cloning, second edition, Cold Spring Harbor Laboratory, Plainview, NY. Typically, a probe is at least about 20 nucleotides in

length. For example, a probe corresponding to a 20 nucleotide sequence set forth in SEQ ID NO: 1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, or 142 can be used to identify an identical or similar nucleic acid. In addition, probes longer or shorter than 20 nucleotides can be used.

The invention also provides isolated nucleic acid sequences that are at least about 12 bases in length (e.g., at least about 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, 60, 100, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, or 5000 bases in length) and hybridize, under hybridization conditions, to the sense or antisense strand of a nucleic acid having the sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163. The hybridization conditions can be moderately or highly stringent hybridization conditions.

For the purpose of this invention, moderately stringent hybridization conditions mean the hybridization is performed at about 42°C in a hybridization solution containing 25 mM $KPO_4$ (pH 7.4), 5X SSC, 5X Denhart's solution, 50 µg/mL denatured, sonicated salmon sperm DNA, 50% formamide, 10% Dextran sulfate, and 1-15 ng/mL probe (about $5x10^7$ cpm/µg), while the washes are performed at about 50°C with a wash solution containing 2X SSC and 0.1% sodium dodecyl sulfate.

Highly stringent hybridization conditions mean the hybridization is performed at about 42°C in a hybridization solution containing 25 mM $KPO_4$ (pH 7.4), 5X SSC, 5X Denhart's solution, 50 µg/mL denatured, sonicated salmon sperm DNA, 50% formamide, 10% Dextran sulfate, and 1-15 ng/mL probe (about $5x10^7$ cpm/µg), while the washes are performed at about 65°C with a wash solution containing 0.2X SSC and 0.1% sodium dodecyl sulfate.

**Purified:** The term "purified" as used herein does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified polypeptide or nucleic acid preparation can be one in which the subject polypeptide or nucleic acid, respectively, is at a higher concentration than the polypeptide or nucleic acid would be in its natural environment within an organism. For example, a polypeptide preparation can be considered purified if the polypeptide content in the preparation represents at least 50%, 60%, 70%, 80%, 85%, 90%, 92%, 95%, 98%, or 99% of the total protein content of the preparation.

**Transformed:** A "transformed" cell is a cell into which a nucleic acid molecule has been introduced by, for example, molecular biology techniques. As used herein, the term "transformation" encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell including, without limitation, transfection with a viral vector, conjugation, transformation with a plasmid vector, and introduction of naked DNA by electroporation, lipofection, and particle gun acceleration.

**Recombinant:** A "recombinant" nucleic acid is one having (1) a sequence that is not naturally occurring in the organism in which it is expressed or (2) a sequence made by an artificial combination of two otherwise-separated, shorter sequences. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. "Recombinant" is also used to describe nucleic acid molecules that have been artificially manipulated, but contain the same regulatory sequences and coding regions that are found in the organism from which the nucleic acid was isolated.

**Specific binding agent:** A "specific binding agent" is an agent that is capable of specifically binding to any of the polypeptide described herein, and can include polyclonal antibodies, monoclonal antibodies (including humanized monoclonal antibodies), and fragments of monoclonal antibodies such as Fab, $F(ab')_2$, and Fv fragments as well as any other agent capable of specifically binding to an epitope of such polypeptides.

Antibodies to the polypeptides provided herein (or fragments thereof) can be used to purify or identify such polypeptides. The amino acid and nucleic acid sequences provided herein allow for the production of specific antibody-based binding agents that recognize the polypeptides described herein.

Monoclonal or polyclonal antibodies can be produced to the polypeptides, portions of the polypeptides, or variants thereof. Optimally, antibodies raised against one or more epitopes on a polypeptide antigen will specifically detect that polypeptide. That is, antibodies raised against one particular polypeptide would recognize and bind that particular polypeptide, and would not substantially recognize or bind to other polypeptides. The determination that an antibody specifically binds to a particular polypeptide is made by any one of a number of standard immunoassay methods; for

instance, Western blotting (See, e.g., Sambrook *et al.* (ed.), Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

To determine that a given antibody preparation (such as a preparation produced in a mouse against a polypeptide having the amino acid sequence set forth in SEQ ID NO: 2) specifically detects the appropriate polypeptide (e.g., a polypeptide having the amino acid sequence set forth in SEQ ID NO: 2) by Western blotting, total cellular protein can be extracted from cells and separated by SDS-polyacrylamide gel electrophoresis. The separated total cellular protein can then be transferred to a membrane (e.g., nitrocellulose), and the antibody preparation incubated with the membrane. After washing the membrane to remove non-specifically bound antibodies, the presence of specifically bound antibodies can be detected using an appropriate secondary antibody (e.g., an anti-mouse antibody) conjugated to an enzyme such as alkaline phosphatase since application of 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium results in the production of a densely blue-colored compound by immuno-localized alkaline phosphatase.

Substantially pure polypeptides suitable for use as an immunogen can be obtained from transfected cells, transformed cells, or wild-type cells. Polypeptide concentrations in the final preparation can be adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms per milliliter. In addition, polypeptides ranging in size from full-length polypeptides to polypeptides having as few as nine amino acid residues can be utilized as immunogens. Such polypeptides can be produced in cell culture, can be chemically synthesized using standard methods, or can be obtained by cleaving large polypeptides into smaller polypeptides that can be purified. Polypeptides having as few as nine amino acid residues in length can be immunogenic when presented to an immune system in the context of a Major Histocompatibility Complex (MHC) molecule such as an MHC class I or MHC class II molecule. Accordingly, polypeptides having at least 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 900, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, or more consecutive amino acid residues of any

amino acid sequence disclosed herein can be used as immunogens for producing antibodies.

Monoclonal antibodies to any of the polypeptides disclosed herein can be prepared from murine hybridomas according to the classic method of Kohler & Milstein (Nature 256:495 (1975)) or a derivative method thereof.

Polyclonal antiserum containing antibodies to the heterogeneous epitopes of any polypeptide disclosed herein can be prepared by immunizing suitable animals with the polypeptide (or fragment thereof), which can be unmodified or modified to enhance immunogenicity. An effective immunization protocol for rabbits can be found in Vaitukaitis *et al.* (*J. Clin. Endocrinol. Metab.* 33:988-991 (1971)).

Antibody fragments can be used in place of whole antibodies and can be readily expressed in prokaryotic host cells. Methods of making and using immunologically effective portions of monoclonal antibodies, also referred to as "antibody fragments," are well known and include those described in Better & Horowitz (*Methods Enzymol.* 178:476-496 (1989)), Glockshuber *et al.* (*Biochemistry* 29:1362-1367 (1990), U.S. Pat. No. 5,648,237 ("Expression of Functional Antibody Fragments"), U.S. Pat. No. 4,946,778 ("Single Polypeptide Chain Binding Molecules"), U.S. Pat. No. 5,455,030 ("Immunotherapy Using Single Chain Polypeptide Binding Molecules"), and references cited therein.

**Operably linked:** A first nucleic acid sequence is "operably linked" with a second nucleic acid sequence whenever the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two polypeptide-coding regions, in the same reading frame.

**Probes and primers:** Nucleic acid probes and primers can be prepared readily based on the amino acid sequences and nucleic acid sequences provided herein. A "probe" includes an isolated nucleic acid containing a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed in, for example, Sambrook *et al.* (ed.), Molecular Cloning:

A Laboratory Manual 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, and Ausubel *et al.* (ed.) Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, New York (with periodic updates), 1987.

"Primers" are typically nucleic acid molecules having ten or more nucleotides (e.g., nucleic acid molecules having between about 10 nucleotides and about 100 nucleotides). A primer can be annealed to a complementary target nucleic acid strand by nucleic acid hybridization to form a hybrid between the primer and the target nucleic acid strand, and then extended along the target nucleic acid strand by, for example, a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, for example, by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

Methods for preparing and using probes and primers are described, for example, in references such as Sambrook *et al.* (ed.), Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989; Ausubel *et al.* (ed.), Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, New York (with periodic updates), 1987; and Innis *et al.*, PCR Protocols: A Guide to Methods and Applications, Academic Press: San Diego, 1990. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, .COPYRGT. 1991, Whitehead Institute for Biomedical Research, Cambridge, Mass.). One of skill in the art will appreciate that the specificity of a particular probe or primer increases with the length, but that a probe or primer can range in size from a full-length sequence to sequences as short as five consecutive nucleotides. Thus, for example, a primer of 20 consecutive nucleotides can anneal to a target with a higher specificity than a corresponding primer of only 15 nucleotides. Thus, in order to obtain greater specificity, probes and primers can be selected that comprise, for example, 10, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 3000, 3050,

3100, 3150, 3200, 3250, 3300, 3350, 3400, 3450, 3500, 3550, 3600, 3650, 3700, 3750, 3800, 3850, 3900, 4000, 4050, 4100, 4150, 4200, 4250, 4300, 4350, 4400, 4450, 4500, 4550, 4600, 4650, 4700, 4750, 4800, 4850, 4900, 5000, 5050, 5100, 5150, 5200, 5250, 5300, 5350, 5400, 5450, or more consecutive nucleotides.

**Percent sequence identity:** The "percent sequence identity" between a particular nucleic acid or amino acid sequence and a sequence referenced by a particular sequence identification number is determined as follows. First, a nucleic acid or amino acid sequence is compared to the sequence set forth in a particular sequence identification number using the BLAST 2 Sequences (Bl2seq) program from the stand-alone version of BLASTZ containing BLASTN version 2.0.14 and BLASTP version 2.0.14. This stand-alone version of BLASTZ can be obtained from Fish & Richardson's web site (www.fr.com) or the United States government's National Center for Biotechnology Information web site (www.ncbi.nlm.nih.gov). Instructions explaining how to use the Bl2seq program can be found in the readme file accompanying BLASTZ. Bl2seq performs a comparison between two sequences using either the BLASTN or BLASTP algorithm. BLASTN is used to compare nucleic acid sequences, while BLASTP is used to compare amino acid sequences. To compare two nucleic acid sequences, the options are set as follows: -i is set to a file containing the first nucleic acid sequence to be compared (e.g., C:\seq1.txt); -j is set to a file containing the second nucleic acid sequence to be compared (e.g., C:\seq2.txt); -p is set to blastn; -o is set to any desired file name (e.g., C:\output.txt); -q is set to –1; -r is set to 2; and all other options are left at their default setting. For example, the following command can be used to generate an output file containing a comparison between two sequences: C:\Bl2seq –i c:\seq1.txt –j c:\seq2.txt –p blastn –o c:\output.txt –q –1 –r 2. To compare two amino acid sequences, the options of Bl2seq are set as follows: -i is set to a file containing the first amino acid sequence to be compared (e.g., C:\seq1.txt); -j is set to a file containing the second amino acid sequence to be compared (e.g., C:\seq2.txt); -p is set to blastp; -o is set to any desired file name (e.g., C:\output.txt); and all other options are left at their default setting. For example, the following command can be used to generate an output file containing a comparison between two amino acid sequences: C:\Bl2seq –i c:\seq1.txt –j c:\seq2.txt –p blastp –o c:\output.txt. If the two compared sequences share homology, then the

designated output file will present those regions of homology as aligned sequences. If the two compared sequences do not share homology, then the designated output file will not present aligned sequences.

Once aligned, the number of matches is determined by counting the number of positions where an identical nucleotide or amino acid residue is presented in both sequences. The percent sequence identity is determined by dividing the number of matches either by the length of the sequence set forth in the identified sequence (e.g., SEQ ID NO:1), or by an articulated length (e.g., 100 consecutive nucleotides or amino acid residues from a sequence set forth in an identified sequence), followed by multiplying the resulting value by 100. For example, a nucleic acid sequence that has 1166 matches when aligned with the sequence set forth in SEQ ID NO:1 is 75.0 percent identical to the sequence set forth in SEQ ID NO:1 (i.e., 1166÷1554*100=75.0). It is noted that the percent sequence identity value is rounded to the nearest tenth. For example, 75.11, 75.12, 75.13, and 75.14 is rounded down to 75.1, while 75.15, 75.16, 75.17, 75.18, and 75.19 is rounded up to 75.2. It is also noted that the length value will always be an integer. In another example, a target sequence containing a 20-nucleotide region that aligns with 20 consecutive nucleotides from an identified sequence as follows contains a region that shares 75 percent sequence identity to that identified sequence (i.e., 15÷20*100=75).

```
                          1                   20
Target Sequence:          AGGTCGTGTACTGTCAGTCA
                          | ||  |||  ||||  ||||  |
Identified Sequence:      ACGTGGTGAACTGCCAGTGA
```

**Conservative substitution:** The term "conservative substitution" as used herein refers to any of the amino acid substitutions set forth in Table 1. Typically, conservative substitutions have little to no impact on the activity of a polypeptide. A polypeptide can be produced to contain one or more conservative substitutions by manipulating the nucleotide sequence that encodes that polypeptide using, for example, standard procedures such as site-directed mutagenesis or PCR.

Table 1

| Original Residue | Conservative Substitution(s) |
|---|---|
| Ala | ser |
| Arg | lys |
| Asn | gln; his |
| Asp | glu |
| Cys | ser |
| Gln | asn |
| Glu | asp |
| Gly | pro |
| His | asn; gln |
| Ile | leu; val |
| Leu | ile; val |
| Lys | arg; gln; glu |
| Met | leu; ile |
| Phe | met; leu; tyr |
| Ser | thr |
| Thr | ser |
| Trp | tyr |
| Tyr | trp; phe |
| Val | ile; leu |

## II. Metabolic Pathways

The invention provides methods and materials related to producing 3-HP as well as other organic compounds (e.g., 1,3-propanediol, acrylic acid, polymerized acrylate, esters of acrylate, polymerized 3-HP, and esters of 3-HP). Specifically, the invention provides isolated nucleic acids, polypeptides, host cells, and methods and materials for producing 3-HP as well as other organic compounds such as 1,3-propanediol, acrylic acid, polymerized acrylate, esters of acrylate, polymerized 3-HP, and esters of 3-HP.

Accordingly, the invention provides several metabolic pathways that can be used to produce organic compounds from PEP (Figures 1-5, 43-44, 54, and 55). As depicted in Figure 1, lactate can be converted into lactyl-CoA by a polypeptide having CoA transferase activity (EC 2.8.3.1); the resulting lactyl-CoA can be converted into acrylyl-CoA by a polypeptide (or multiple polypeptide complex such as an activated E2 α and E2 β complex) having lactyl-CoA dehydratase activity (EC 4.2.1.54); the resulting acrylyl-CoA can be converted into 3-hydroxypropionyl-CoA (3-HP-CoA) by a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity (EC 4.2.1.-); and the resulting 3-HP-CoA can be converted into 3-HP by a polypeptide having CoA transferase activity, a polypeptide having 3-hydroxypropionyl-CoA hydrolase activity (EC 3.1.2.-), or a polypeptide having 3-hydroxyisobutryl-CoA hydrolase activity (EC 3.1.2.4).

Polypeptides having CoA transferase activity as well as nucleic acid encoding such polypeptides can be obtained from various species including, without limitation, *Megasphaera elsdenii*, *Clostridium propionicum*, *Clostridium kluyveri*, and *Escherichia coli*. For example, nucleic acid that encodes a polypeptide having CoA transferase activity can be obtained from *Megasphaera elsdenii* as described in Example 1 and can have a sequence as set forth in SEQ ID NO: 1. In addition, polypeptides having CoA transferase activity as well as nucleic acid encoding such polypeptides can be obtained as described herein. For example, the variations to SEQ ID NO: 1 provided herein can be used to encode a polypeptide having CoA transferase activity.

Polypeptides (or the polypeptides of a multiple polypeptide complex such as an activated E2 α and E2 β complex) having lactyl-CoA dehydratase activity as well as nucleic acid encoding such polypeptides can be obtained from various species including, without limitation, *Megasphaera elsdenii* and *Clostridium propionicum*. For example, nucleic acid encoding an E1 activator, an E2 α subunit, and an E2 β subunit that can form a multiple polypeptide complex having lactyl-CoA dehydratase activity can be obtained from *Megasphaera elsdenii* as described in Example 2. The nucleic acid encoding the E1 activator can contain a sequence as set forth in SEQ ID NO: 9; the nucleic acid encoding the E2 α subunit can contain a sequence as set forth in SEQ ID NO: 17; and the nucleic acid encoding the E2 β subunit can contain a sequence as set forth in SEQ ID NO: 25. In addition, polypeptides (or the polypeptides of a multiple polypeptide complex) having

lactyl-CoA dehydratase activity as well as nucleic acid encoding such polypeptides can be obtained as described herein. For example, the variations to SEQ ID NO: 9, 17, and 25 provided herein can be used to encode the polypeptides of a multiple polypeptide complex having CoA transferase activity.

Polypeptides having 3-hydroxypropionyl-CoA dehydratase activity as well as nucleic acid encoding such polypeptides can be obtained from various species including, without limitation, *Chloroflexus aurantiacus*, *Candida rugosa*, *Rhodosprillium rubrum*, and *Rhodobacter capsulates*. For example, nucleic acid that encodes a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity can be obtained from *Chloroflexus aurantiacus* as described in Example 3 and can have a sequence as set forth in SEQ ID NO: 40. In addition, polypeptides having 3-hydroxypropionyl-CoA dehydratase activity as well as nucleic acid encoding such polypeptides can be obtained as described herein. For example, the variations to SEQ ID NO: 40 provided herein can be used to encode a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity.

Polypeptides having 3-hydroxypropionyl-CoA hydrolase activity as well as nucleic acid encoding such polypeptides can be obtained from various species including, without limitation, *Candida rugosa*. Polypeptides having 3-hydroxyisobutryl-CoA hydrolase activity as well as nucleic acid encoding such polypeptides can be obtained from various species including, without limitation, *Pseudomonas fluorescens, rattus,* and *homo sapiens*. For example, nucleic acid that encodes a polypeptide having 3-hydroxyisobutryl-CoA hydrolase activity can be obtained from *homo sapiens* and can have a sequence as set forth in GenBank® accession number U66669.

The term "polypeptide having enzymatic activity" as used herein refers to any polypeptide that catalyzes a chemical reaction of other substances without itself being destroyed or altered upon completion of the reaction. Typically, a polypeptide having enzymatic activity catalyzes the formation of one or more products from one or more substrates. Such polypeptides can have any type of enzymatic activity including, without limitation, the enzymatic activity or enzymatic activities associated with enzymes such as dehydratases/hydratases, 3-hydroxypropionyl-CoA dehydratases/hydratases, CoA transferases, lactyl-CoA dehydratases, 3-hydroxypropionyl-CoA hydrolases, 3-

hydroxyisobutryl-CoA hydrolases, poly hydroxyacid synthases, CoA synthetases, malonyl-CoA reductases, β-alanine ammonia lyases, and lipases.

As depicted in Figure 2, lactate can be converted into lactyl-CoA by a polypeptide having CoA synthetase activity (EC 6.2.1.-); the resulting lactyl-CoA can be converted into acrylyl-CoA by a polypeptide (or multiple polypeptide complex) having lactyl-CoA dehydratase activity; the resulting acrylyl-CoA can be converted into 3-HP-CoA by a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity; and the resulting 3-HP-CoA can be converted into polymerized 3-HP by a polypeptide having poly hydroxyacid synthase activity (EC 2.3.1.-). Polypeptides having CoA synthetase activity as well as nucleic acid encoding such polypeptides can be obtained from various species including, without limitation, *Escherichia coli*, *Rhodobacter sphaeroides*, *Saccharomyces cervisiae*, and *Salmonella enterica*. For example, nucleic acid that encodes a polypeptide having CoA synthetase activity can be obtained from *Escherichia coli* and can have a sequence as set forth in GenBank® accession number U00006. Polypeptides (or multiple polypeptide complexes) having lactyl-CoA dehydratase activity as well as nucleic acid encoding such polypeptides can be obtained as provided herein. Polypeptides having 3-hydroxypropionyl-CoA dehydratase activity as well as nucleic acid encoding such polypeptides also can be obtained as provided herein. Polypeptides having poly hydroxyacid synthase activity as well as nucleic acid encoding such polypeptides can be obtained from various species including, without limitation, *Rhodobacter sphaeroides*, *Comamonas acidororans*, *Ralstonia eutropha*, and *Pseudomonas oleovorans*. For example, nucleic acid that encodes a polypeptide having poly hydroxyacid synthase activity can be obtained from *Rhodobacter sphaeroides* and can have a sequence as set forth in GenBank® accession number X97200.

As depicted in Figure 3, lactate can be converted into lactyl-CoA by a polypeptide having CoA transferase activity; the resulting lactyl-CoA can be converted into acrylyl-CoA by a polypeptide (or multiple polypeptide complex) having lactyl-CoA dehydratase activity; the resulting acrylyl-CoA can be converted into 3-HP-CoA by a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity; the resulting 3-HP-CoA can be converted into 3-HP by a polypeptide having CoA transferase activity, a polypeptide having 3-hydroxypropionyl-CoA hydrolase activity, or a polypeptide having 3-

hydroxyisobutryl-CoA hydrolase activity; and the resulting 3-HP can be converted into an ester of 3-HP by a polypeptide having lipase activity (EC 3.1.1.-). Polypeptides having lipase activity as well as nucleic acid encoding such polypeptides can be obtained from various species including, without limitation, *Candida rugosa*, *Candida tropicalis*, and *Candida albicans*. For example, nucleic acid that encodes a polypeptide having lipase activity can be obtained from *Candida rugosa* and can have a sequence as set forth in GenBank® accession number A81171.

As depicted in Figure 4, lactate can be converted into lactyl-CoA by a polypeptide having CoA synthetase activity; the resulting lactyl-CoA can be converted into acrylyl-CoA by a polypeptide (or multiple polypeptide complex) having lactyl-CoA dehydratase activity; and the resulting acrylyl-CoA can be converted into polymerized acrylate by a polypeptide having poly hydroxyacid synthase activity.

As depicted in Figure 5, lactate can be converted into lactyl-CoA by a polypeptide having CoA transferase activity; the resulting lactyl-CoA can be converted into acrylyl-CoA by a polypeptide (or multiple polypeptide complex) having lactyl-CoA dehydratase activity; the resulting acrylyl-CoA can be converted into acrylate by a polypeptide having CoA transferase activity; and the resulting acrylate can be converted into an ester of acrylate by a polypeptide having lipase activity.

As depicted in Figure 44, acetyl-CoA can be converted into malonyl-CoA by a polypeptide having acetyl-CoA carboxylase activity, and the resulting malonyl-CoA can be converted into 3-HP by a polypeptide having malonyl-CoA reductase activity. Polypeptides having acetyl-CoA carboxylase activity as well as nucleic acid encoding such polypeptides can be obtained from various species including, without limitation, *Escherichia coli* and *Chloroflexus aurantiacus*. For example, nucleic acid that encodes a polypeptide having acetyl-CoA carboxylase activity can be obtained from *Escherichia coli* and can have a sequence as set forth in GenBank® accession number M96394 or U18997. Polypeptides having malonyl-CoA reductase activity as well as nucleic acid encoding such polypeptides can be obtained from various species including, without limitation, *Chloroflexus aurantiacus*, *Sulfolobus metacillus*, and *Acidianus brierleyi*. For example, nucleic acid that encodes a polypeptide having malonyl-CoA reductase activity can be obtained as described herein and can have a sequence similar to the sequence set

forth in SEQ ID NO: 140. In addition, polypeptides having malonyl-CoA reductase activity as well as nucleic acid encoding such polypeptides can be obtained as described herein. For example, the variations to SEQ ID NO: 140 provided herein can be used to encode a polypeptide having malonyl-CoA reductase activity.

Polypeptides having malonyl-CoA reductase activity can use NADPH as a co-factor. For example, the polypeptide having the amino acid sequence set forth in SEQ ID NO: 141 is a polypeptide having malonyl-CoA reductase activity that uses NADPH as a co-factor when converting malonyl-CoA into 3-HP. Likewise, polypeptides having malonyl-CoA reductase activity can use NADH as a co-factor. Such polypeptides can be obtained by converting a polypeptide that has malonyl-CoA reductase activity and uses NADPH as a cofactor into a polypeptide that has malonyl-CoA reductase activity and uses NADH as a cofactor. Any method can be used to convert a polypeptide that uses NADPH as a cofactor into a polypeptide that uses NADH as a cofactor such as those described by others (Eppink *et al.*, *J. Mol. Biol.*, 292(1):87-96 (1999), Hall and Tomsett, *Microbiology*, 146(Pt 6):1399-406 (2000), and Dohr *et al.*, *Proc. Natl. Acad. Sci.*, 98(1):81-86 (2001)). For example, mutagenesis can be used to convert the polypeptide encoded by the nucleic acid sequence set forth in SEQ ID NO: 140 into a polypeptide that, when converting malonyl-CoA into 3-HP, uses NADH as a co-factor instead of NADPH.

As depicted in Figure 43, propionate can be converted into propionyl-CoA by a polypeptide having CoA synthetase activity such as the polypeptide having the sequence set forth in SEQ ID NO: 39; the resulting propionyl-CoA can be converted into acrylyl-CoA by a polypeptide having dehydrogenase activity such as the polypeptide having the sequence set forth in SEQ ID NO: 39; and the resulting acrylyl-CoA can be converted into (1) acrylate by a polypeptide having CoA transferase activity or CoA hydrolase activity, (2) 3-HP-CoA by a polypeptide having 3-HP dehydratase activity (also referred to as acrylyl-CoA hydratase or simply hydratase) such as the polypeptide having the sequence set forth in SEQ ID NO:39, or (3) polymerized acrylate by a polypeptide having poly hydroxyacid synthase activity. The resulting acrylate can be converted into an ester of acrylate by a polypeptide having lipase activity. The resulting 3-HP-CoA can be converted into (1) 3-HP by a polypeptide having CoA transferase activity, a polypeptide

having 3-hydroxypropionyl-CoA hydrolase activity (EC 3.1.2.-), or a polypeptide having 3-hydroxyisobutyryl-CoA hydrolase activity (EC 3.1.2.4), or (2) polymerized 3-HP by a polypeptide having poly hydroxyacid synthase activity (EC 2.3.1.-).

As depicted in Figure 54, PEP can be converted into β-alanine. β-alanine can be converted into β-alanyl-CoA through the use of a polypeptide having CoA transferase activity. β-alanyl-CoA can then be converted into acrylyl-CoA through the use of a polypeptide having β-alanyl-CoA ammonia lyase activity. Acrylyl-CoA can then be converted into 3-HP-CoA through the use of a polypeptide having 3-HP-CoA dehydratase activity, and a polypeptide having glutamate dehydrogenase activity can be used to convert 3-HP-CoA into 3-HP.

As depicted in Figure 55, 3-HP can be made from β-alanine by first contacting β-alanine with a polypeptide having 4,4-aminobutyrate aminotransferase activity to create malonate semialdehyde. The malonate semialdehyde can be converted into 3-HP with a polypeptide having 3-HP dehydrogenase activity or a polypeptide having 3-hydroxyisobutyrate dehydrogenase activity.

**III. Nucleic acid molecules and polypeptides**

The invention provides isolated nucleic acid that contains the entire nucleic acid sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163. In addition, the invention provides isolated nucleic acid that contains a portion of the nucleic acid sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163. For example, the invention provides isolated nucleic acid that contains a 15 nucleotide sequence identical to any 15 nucleotide sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163 including, without limitation, the sequence starting at nucleotide number 1 and ending at nucleotide number 15, the sequence starting at nucleotide number 2 and ending at nucleotide number 16, the sequence starting at nucleotide number 3 and ending at nucleotide number 17, and so forth. It will be appreciated that the invention also provides isolated nucleic acid that contains a nucleotide sequence that is greater than 15 nucleotides (e.g., 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more nucleotides) in length and identical to any portion of the sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129,

140, 142, 162, or 163. For example, the invention provides isolated nucleic acid that contains a 25 nucleotide sequence identical to any 25 nucleotide sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163 including, without limitation, the sequence starting at nucleotide number 1 and ending at nucleotide number 25, the sequence starting at nucleotide number 2 and ending at nucleotide number 26, the sequence starting at nucleotide number 3 and ending at nucleotide number 27, and so forth. Additional examples include, without limitation, isolated nucleic acids that contain a nucleotide sequence that is 50 or more nucleotides (e.g., 100, 150, 200, 250, 300, or more nucleotides) in length and identical to any portion of the sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163. Such isolated nucleic acids can include, without limitation, those isolated nucleic acids containing a nucleic acid sequence represented in a single line of sequence depicted in Figure 6, 10, 14, 18, 22, 23, 25, 27, 29, 31, 39, 49, or 51 since each line of sequence depicted in these figures, with the possible exception of the last line, provides a nucleotide sequence containing at least 50 bases.

In addition, the invention provides isolated nucleic acid that contains a variation of the nucleic acid sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163. For example, the invention provides isolated nucleic acid containing a nucleic acid sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163 that contains a single insertion, a single deletion, a single substitution, multiple insertions, multiple deletions, multiple substitutions, or any combination thereof (e.g., single deletion together with multiple insertions). Such isolated nucleic acid molecules can share at least 60, 65, 70, 75, 80, 85, 90, 95, 97, 98, or 99 percent sequence identity with a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163.

The invention provides multiple examples of isolated nucleic acid that contains a variation of a nucleic acid sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163. For example, Figure 8 provides the sequence set forth in SEQ ID NO:1 aligned with three other nucleic acid sequences. Examples of variations of the sequence set forth in SEQ ID NO:1 include, without limitation, any variation of the sequence set forth in SEQ ID NO:1 provided in Figure 8. Such variations are provided in

Figure 8 in that a comparison of the nucleotide (or lack thereof) at a particular position of the sequence set forth in SEQ ID NO:1 with the nucleotide (or lack thereof) at the same aligned position of any of the other three nucleic acid sequences depicted in Figure 8 (i.e., SEQ ID NOs:3, 4, and 5) provides a list of specific changes for the sequence set forth in SEQ ID NO:1. For example, the "a" at position 49 of SEQ ID NO:1 can be substituted with an "c" as indicated in Figure 8. As also indicated in Figure 8, the "a" at position 590 of SEQ ID NO:1 can be substituted with a "atgg"; an "aaac" can be inserted before the "g" at position 393 of SEQ ID NO:1; or the "gaa" at position 736 of SEQ ID NO:1 can be deleted. It will be appreciated that the sequence set forth in SEQ ID NO:1 can contain any number of variations as well as any combination of types of variations. For example, the sequence set forth in SEQ ID NO:1 can contain one variation provided in Figure 8 or more than one (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, or more) of the variations provided in Figure 8. It is noted that the nucleic acid sequences provided by Figure 8 can encode polypeptides having CoA transferase activity. The invention also provides isolated nucleic acid that contains a variant of a portion of the sequence set forth in SEQ ID NO:1 as depicted in Figure 8 and described herein.

Likewise, Figure 12 provides variations of SEQ ID NO:9 and portions thereof; Figure 16 provides variations of SEQ ID NO:17 and portions thereof; Figure 20 provides variations of SEQ ID NO:25 and portions thereof; Figure 32 provides variations of SEQ ID NO:40 and portions thereof; and Figure 53 provides variations of SEQ ID NO:140.

The invention provides isolated nucleic acid that contains a nucleic acid sequence that encodes the entire amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. In addition, the invention provides isolated nucleic acid that contains a nucleic acid sequence that encodes a portion of the amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. For example, the invention provides isolated nucleic acid that contains a nucleic acid sequence that encodes a 15 amino acid sequence identical to any 15 amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161 including, without limitation, the sequence starting at amino acid residue number 1 and ending at amino acid residue number 15, the sequence starting at amino acid residue number 2 and ending at amino acid residue number 16, the sequence starting at amino acid residue number 3 and ending

at amino acid residue number 17, and so forth. It will be appreciated that the invention also provides isolated nucleic acid that contains a nucleic acid sequence that encodes an amino acid sequence that is greater than 15 amino acid residues (e.g., 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more amino acid residues) in length and identical to any portion of the sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. For example, the invention provides isolated nucleic acid that contains a nucleic acid sequence that encodes a 25 amino acid sequence identical to any 25 amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161 including, without limitation, the sequence starting at amino acid residue number 1 and ending at amino acid residue number 25, the sequence starting at amino acid residue number 2 and ending at amino acid residue number 26, the sequence starting at amino acid residue number 3 and ending at amino acid residue number 27, and so forth. Additional examples include, without limitation, isolated nucleic acids that contain a nucleic acid sequence that encodes an amino acid sequence that is 50 or more amino acid residues (e.g., 100, 150, 200, 250, 300, or more amino acid residues) in length and identical to any portion of the sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. Such isolated nucleic acids can include, without limitation, those isolated nucleic acids containing a nucleic acid sequence that encodes an amino acid sequence represented in a single line of sequence depicted in Figure 7, 11, 15, 19, 24, 26, 28, 30, or 50 since each line of sequence depicted in these figures, with the possible exception of the last line, provides an amino acid sequence containing at least 50 residues.

In addition, the invention provides isolated nucleic acid that contains a nucleic acid sequence that encodes an amino acid sequence having a variation of the amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. For example, the invention provides isolated nucleic acid containing a nucleic acid sequence encoding an amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161 that contains a single insertion, a single deletion, a single substitution, multiple insertions, multiple deletions, multiple substitutions, or any combination thereof (e.g., single deletion together with multiple insertions). Such isolated nucleic acid molecules can contain a nucleic acid sequence encoding an amino acid sequence that shares at least 60, 65, 70, 75, 80, 85, 90, 95, 97, 98, or 99 percent sequence identity with a

sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161.

The invention provides multiple examples of isolated nucleic acid containing a nucleic acid sequence encoding an amino acid sequence having a variation of an amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. For example, Figure 9 provides the amino acid sequence set forth in SEQ ID NO:2 aligned with three other amino acid sequences. Examples of variations of the sequence set forth in SEQ ID NO:2 include, without limitation, any variation of the sequence set forth in SEQ ID NO:2 provided in Figure 9. Such variations are provided in Figure 9 in that a comparison of the amino acid residue (or lack thereof) at a particular position of the sequence set forth in SEQ ID NO:2 with the amino acid residue (or lack thereof) at the same aligned position of any of the other three amino acid sequences of Figure 9 (i.e., SEQ ID NOs:6, 7, and 8) provides a list of specific changes for the sequence set forth in SEQ ID NO:2. For example, the "k" at position 17 of SEQ ID NO:2 can be substituted with a "p" or "h" as indicated in Figure 9. As also indicated in Figure 9, the "v" at position 125 of SEQ ID NO:2 can be substituted with an "i" or "f". It will be appreciated that the sequence set forth in SEQ ID NO:2 can contain any number of variations as well as any combination of types of variations. For example, the sequence set forth in SEQ ID NO:2 can contain one variation provided in Figure 9 or more than one (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, or more) of the variations provided in Figure 9. It is noted that the amino acid sequences provided in Figure 9 can be polypeptides having CoA transferase activity.

The invention also provides isolated nucleic acid containing a nucleic acid sequence encoding an amino acid sequence that contains a variant of a portion of the sequence set forth in SEQ ID NO:2 as depicted in Figure 9 and described herein.

Likewise, Figure 13 provides variations of SEQ ID NO:10 and portions thereof; Figure 17 provides variations of SEQ ID NO:18 and portions thereof; Figure 21 provides variations of SEQ ID NO:26 and portions thereof; Figure 33 provides variations of SEQ ID NO:41 and portions thereof; Figures 40, 41, and 42 provide variations of SEQ ID NO:39; and Figure 52 provides variations of SEQ ID NO:141 and portions thereof.

It is noted that codon preferences and codon usage tables for a particular species can be used to engineer isolated nucleic acid molecules that take advantage of the codon

usage preferences of that particular species. For example, the isolated nucleic acid provided herein can be designed to have codons that are preferentially used by a particular organism of interest.

The invention also provides isolated nucleic acid that is at least about 12 bases in length (e.g., at least about 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, 60, 100, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, or 5000 bases in length) and hybridizes, under hybridization conditions, to the sense or antisense strand of a nucleic acid having the sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163. The hybridization conditions can be moderately or highly stringent hybridization conditions.

The invention provides polypeptides that contain the entire amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. In addition, the invention provides polypeptides that contain a portion of the amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. For example, the invention provides polypeptides that contain a 15 amino acid sequence identical to any 15 amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161 including, without limitation, the sequence starting at amino acid residue number 1 and ending at amino acid residue number 15, the sequence starting at amino acid residue number 2 and ending at amino acid residue number 16, the sequence starting at amino acid residue number 3 and ending at amino acid residue number 17, and so forth. It will be appreciated that the invention also provides polypeptides that contain an amino acid sequence that is greater than 15 amino acid residues (e.g., 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more amino acid residues) in length and identical to any portion of the sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. For example, the invention provides polypeptides that contain a 25 amino acid sequence identical to any 25 amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161 including, without limitation, the sequence starting at amino acid residue number 1 and ending at amino acid residue number 25, the sequence starting at amino acid residue number 2 and ending at amino acid residue number 26, the sequence starting at amino acid residue number 3 and ending at amino acid residue number 27, and so forth. Additional examples include, without limitation, polypeptides

that contain an amino acid sequence that is 50 or more amino acid residues (e.g., 100, 150, 200, 250, 300, or more amino acid residues) in length and identical to any portion of the sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. Such polypeptides can include, without limitation, those polypeptides containing a amino acid sequence represented in a single line of sequence depicted in Figure 7, 11, 15, 19, 24, 26, 28, 30, or 50 since each line of sequence depicted in these figures, with the possible exception of the last line, provides an amino acid sequence containing at least 50 residues.

In addition, the invention provides polypeptides that an amino acid sequence having a variation of the amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. For example, the invention provides polypeptides containing an amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161 that contains a single insertion, a single deletion, a single substitution, multiple insertions, multiple deletions, multiple substitutions, or any combination thereof (e.g., single deletion together with multiple insertions). Such polypeptides can contain an amino acid sequence that shares at least 60, 65, 70, 75, 80, 85, 90, 95, 97, 98, or 99 percent sequence identity with a sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161.

The invention provides multiple examples of polypeptides containing an amino acid sequence having a variation of an amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161. For example, Figure 9 provides the amino acid sequence set forth in SEQ ID NO:2 aligned with three other amino acid sequences. Examples of variations of the sequence set forth in SEQ ID NO:2 include, without limitation, any variation of the sequence set forth in SEQ ID NO:2 provided in Figure 9. Such variations are provided in Figure 9 in that a comparison of the amino acid residue (or lack thereof) at a particular position of the sequence set forth in SEQ ID NO:2 with the amino acid residue (or lack thereof) at the same aligned position of any of the other three amino acid sequences of Figure 9 (i.e., SEQ ID NOs:6, 7, and 8) provides a list of specific changes for the sequence set forth in SEQ ID NO:2. For example, the "k" at position 17 of SEQ ID NO:2 can be substituted with a "p" or "h" as indicated in Figure 9. As also indicated in Figure 9, the "v" at position 125 of SEQ ID NO:2 can be substituted with an "i" or "f". It will be appreciated that the sequence set forth in SEQ ID NO:2 can

contain any number of variations as well as any combination of types of variations. For example, the sequence set forth in SEQ ID NO:2 can contain one variation provided in Figure 9 or more than one (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, or more) of the variations provided in Figure 9. It is noted that the amino acid sequences provided in Figure 9 can be polypeptides having CoA transferase activity.

The invention also provides polypeptides containing an amino acid sequence that contains a variant of a portion of the sequence set forth in SEQ ID NO:2 as depicted in Figure 9 and described herein.

Likewise, Figure 13 provides variations of SEQ ID NO:10 and portions thereof; Figure 17 provides variations of SEQ ID NO:18 and portions thereof; Figure 21 provides variations of SEQ ID NO:26 and portions thereof; Figure 33 provides variations of SEQ ID NO:41 and portions thereof, Figures 40, 41, and 42 provide variations of SEQ ID NO:39; and Figure 52 provides variations of SEQ ID NO:141 and portions thereof.

Polypeptides having a variant amino acid sequence can retain enzymatic activity. Such polypeptides can be produced by manipulating the nucleotide sequence encoding a polypeptide using standard procedures such as site-directed mutagenesis or PCR. One type of modification includes the substitution of one or more amino acid residues for amino acid residues having a similar biochemical property. For example, a polypeptide can have an amino acid sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161 with one or more conservative substitutions.

More substantial changes can be obtained by selecting substitutions that are less conservative than those in Table 1, i.e., selecting residues that differ more significantly in their effect on maintaining: (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation; (b) the charge or hydrophobicity of the polypeptide at the target site; or (c) the bulk of the side chain. The substitutions that in general are expected to produce the greatest changes in polypeptide function are those in which: (a) a hydrophilic residue, e.g., serine or threonine, is substituted for (or by) a hydrophobic residue, e.g., leucine, isoleucine, phenylalanine, valine or alanine; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysine, arginine, or histidine, is substituted for (or by) an electronegative residue, e.g., glutamic acid or aspartic acid; or

(d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine. The effects of these amino acid substitutions (or other deletions or additions) can be assessed for polypeptides having enzymatic activity by analyzing the ability of the polypeptide to catalyze the conversion of the same substrate as the related native polypeptide to the same product as the related native polypeptide. Accordingly, polypeptides having 5, 10, 20, 30, 40, 50 or less conservative substitutions are provided by the invention.

Polypeptides and nucleic acid encoding polypeptide can be produced by standard DNA mutagenesis techniques, for example, M13 primer mutagenesis. Details of these techniques are provided in Sambrook *et al.* (ed.), Molecular Cloning: A Laboratory Manual 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring, Harbor, N.Y., 1989, Ch. 15. Nucleic acid molecules can contain changes of a coding region to fit the codon usage bias of the particular organism into which the molecule is to be introduced.

Alternatively, the coding region can be altered by taking advantage of the degeneracy of the genetic code to alter the coding sequence in such a way that, while the nucleic acid sequence is substantially altered, it nevertheless encodes a polypeptide having an amino acid sequence identical or substantially similar to the native amino acid sequence. For example, the ninth amino acid residue of the sequence set forth in SEQ ID NO: 2 is alanine, which is encoded in the open reading frame by the nucleotide codon triplet GCT. Because of the degeneracy of the genetic code, three other nucleotide codon triplets--GCA, GCC, and GCG --also code for alanine. Thus, the nucleic acid sequence of the open reading frame can be changed at this position to any of these three codons without affecting the amino acid sequence of the encoded polypeptide or the characteristics of the polypeptide. Based upon the degeneracy of the genetic code, nucleic acid variants can be derived from a nucleic acid sequence disclosed herein using a standard DNA mutagenesis techniques as described herein, or by synthesis of nucleic acid sequences. Thus, this invention also encompasses nucleic acid molecules that encode the same polypeptide but vary in nucleic acid sequence by virtue of the degeneracy of the genetic code.

## IV. Methods of Making 3-HP and Other Organic Acids

Each step provided in the pathways depicted in Figures 1-5, 43-44, 54, and 55 can be performed within a cell (*in vivo*) or outside a cell (*in vitro*, e.g., in a container or column). Additionally, the organic acid products can be generated through a combination of *in vivo* synthesis and *in vitro* synthesis. Moreover, the *in vitro* synthesis step, or steps, can be via chemical reaction or enzymatic reaction.

For example, a microorganism provided herein can be used to perform the steps provided in Figure 1, or an extract containing polypeptides having the indicated enzymatic activities can be used to perform the steps provided in Figure 1. In addition, chemical treatments can be used to perform the conversions provided in Figures 1-5, 43-44, 54, and 55. For example, acrylyl-CoA can be converted into acrylate by hydrolysis. Other chemical treatments include, without limitation, trans esterification to convert acrylate into an acrylate ester.

Carbon sources suitable as starting points for bioconversion include carbohydrates and synthetic intermediates. Examples of carbohydrates which cells are capable of metabolizing to pyruvate include sugars such as dextrose, triglycerides, and fatty acids.

Additionally, intermediate chemical products can be starting points. For example, acetic acid and carbon dioxide can be introduced into a fermentation broth. Acetyl-CoA, malonyl-CoA, and 3-HP can be sequentially produced using a polypeptide having CoA synthase activity, a polypeptide having acetyl-CoA carboxylase activity, and a polypeptide having malonyl-CoA reductase activity. Other useful intermediate chemical starting points can include propionic acid, acrylic acid, lactic acid, pyruvic acid, and β-alanine.

### A. Expression of Polypeptides

The polypeptides described herein can be produced individually in a host cell or in combination in a host cell. Moreover, the polypeptides having a particular enzymatic activity can be a polypeptide that is either naturally-occurring or non-naturally-occurring. A naturally-occurring polypeptide is any polypeptide having an amino acid sequence as found in nature, including wild-type and polymorphic polypeptides. Such naturally-occurring polypeptides can be obtained from any species including, without limitation,

animal (e.g., mammalian), plant, fungal, and bacterial species. A non-naturally-occurring polypeptide is any polypeptide having an amino acid sequence that is not found in nature. Thus, a non-naturally-occurring polypeptide can be a mutated version of a naturally-occurring polypeptide, or an engineered polypeptide. For example, a non-naturally-occurring polypeptide having 3-hydroxypropionyl-CoA dehydratase activity can be a mutated version of a naturally-occurring polypeptide having 3-hydroxypropionyl-CoA dehydratase activity that retains at least some 3-hydroxypropionyl-CoA dehydratase activity. A polypeptide can be mutated by, for example, sequence additions, deletions, substitutions, or combinations thereof.

The invention provides genetically modified cells that can be used to perform one or more steps of the steps in the metabolic pathways described herein or the genetically modified cells can be used to produce the disclosed polypeptides for subsequent use *in vitro*. For example, an individual microorganism can contain exogenous nucleic acid such that each of the polypeptides necessary to perform the steps depicted in Figures 1, 2, 3, 4, 5, 43, 44, 54, or 55 are expressed. It is important to note that such cells can contain any number of exogenous nucleic acid molecules. For example, a particular cell can contain six exogenous nucleic acid molecules with each one encoding one of the six polypeptides necessary to convert lactate into 3-HP as depicted in Figure 1, or a particular cell can endogenously produce polypeptides necessary to convert lactate into acrylyl-CoA while containing exogenous nucleic acid that encodes polypeptides necessary to convert acrylyl-CoA into 3-HP.

In addition, a single exogenous nucleic acid molecule can encode one or more than one polypeptide. For example, a single exogenous nucleic acid molecule can contain sequences that encode three different polypeptides. Further, the cells described herein can contain a single copy, or multiple copies (e.g., about 5, 10, 20, 35, 50, 75, 100 or 150 copies), of a particular exogenous nucleic acid molecule. For example, a particular cell can contain about 50 copies of the constructs depicted in Figure 34, 35, 36, 37, 38, or 45. Again, the cells described herein can contain more than one particular exogenous nucleic acid molecule. For example, a particular cell can contain about 50 copies of exogenous nucleic acid molecule X as well as about 75 copies of exogenous nucleic acid molecule Y.

In another embodiment, a cell within the scope of the invention can contain an exogenous nucleic acid molecule that encodes a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity. Such cells can have any level of 3-hydroxypropionyl-CoA dehydratase activity. For example, a cell containing an exogenous nucleic acid molecule that encodes a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity can have 3-hydroxypropionyl-CoA dehydratase activity with a specific activity greater than about 1 mg 3-HP-CoA formed per gram dry cell weight per hour (e.g., greater than about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 200, 250, 300, 350, 400, 500, or more mg 3-HP-CoA formed per gram dry cell weight per hour). Alternatively, a cell can have 3-hydroxypropionyl-CoA dehydratase activity such that a cell extract from $1x10^6$ cells has a specific activity greater than about 1 μg 3-HP-CoA formed per mg total protein per 10 minutes (e.g., greater than about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 200, 250, 300, 350, 400, 500, or more μg 3-HP-CoA formed per mg total protein per 10 minutes).

A nucleic acid molecule encoding a polypeptide having enzymatic activity can be identified and obtained using any method such as those described herein. For example, nucleic acid molecules that encode a polypeptide having enzymatic activity can be identified and obtained using common molecular cloning or chemical nucleic acid synthesis procedures and techniques, including PCR. In addition, standard nucleic acid sequencing techniques and software programs that translate nucleic acid sequences into amino acid sequences based on the genetic code can be used to determine whether or not a particular nucleic acid has any sequence homology with known enzymatic polypeptides. Sequence alignment software such as MEGALIGN® (DNASTAR, Madison, WI, 1997) can be used to compare various sequences. In addition, nucleic acid molecules encoding known enzymatic polypeptides can be mutated using common molecular cloning techniques (e.g., site-directed mutagenesis). Possible mutations include, without limitation, deletions, insertions, and base substitutions, as well as combinations of deletions, insertions, and base substitutions. Further, nucleic acid and amino acid databases (e.g., GenBank®) can be used to identify a nucleic acid sequence that encodes a polypeptide having enzymatic activity. Briefly, any amino acid sequence having some homology to a polypeptide having enzymatic activity, or any nucleic acid sequence

having some homology to a sequence encoding a polypeptide having enzymatic activity can be used as a query to search GenBank®. The identified polypeptides then can be analyzed to determine whether or not they exhibit enzymatic activity.

In addition, nucleic acid hybridization techniques can be used to identify and obtain a nucleic acid molecule that encodes a polypeptide having enzymatic activity. Briefly, any nucleic acid molecule that encodes a known enzymatic polypeptide, or fragment thereof, can be used as a probe to identify a similar nucleic acid molecules by hybridization under conditions of moderate to high stringency. Such similar nucleic acid molecules then can be isolated, sequenced, and analyzed to determine whether the encoded polypeptide has enzymatic activity.

Expression cloning techniques also can be used to identify and obtain a nucleic acid molecule that encodes a polypeptide having enzymatic activity. For example, a substrate known to interact with a particular enzymatic polypeptide can be used to screen a phage display library containing that enzymatic polypeptide. Phage display libraries can be generated as described elsewhere (Burritt *et al.*, *Anal. Biochem.* 238:1-13 (1990)), or can be obtained from commercial suppliers such as Novagen (Madison, WI).

Further, polypeptide sequencing techniques can be used to identify and obtain a nucleic acid molecule that encodes a polypeptide having enzymatic activity. For example, a purified polypeptide can be separated by gel electrophoresis, and its amino acid sequence determined by, for example, amino acid microsequencing techniques. Once determined, the amino acid sequence can be used to design degenerate oligonucleotide primers. Degenerate oligonucleotide primers can be used to obtain the nucleic acid encoding the polypeptide by PCR. Once obtained, the nucleic acid can be sequenced, cloned into an appropriate expression vector, and introduced into a microorganism.

Any method can be used to introduce an exogenous nucleic acid molecule into a cell. In fact, many methods for introducing nucleic acid into microorganisms such as bacteria and yeast are well known to those skilled in the art. For example, heat shock, lipofection, electroporation, conjugation, fusion of protoplasts, and biolistic delivery are common methods for introducing nucleic acid into bacteria and yeast cells. See, e.g., Ito

*et al.*, *J. Bacterol.* 153:163-168 (1983); Durrens *et al.*, *Curr. Genet.* 18:7-12 (1990); and Becker and Guarente, *Methods in Enzymology* 194:182-187 (1991).

An exogenous nucleic acid molecule contained within a particular cell of the invention can be maintained within that cell in any form. For example, exogenous nucleic acid molecules can be integrated into the genome of the cell or maintained in an episomal state. In other words, a cell of the invention can be a stable or transient transformant. Again, a microorganism described herein can contain a single copy, or multiple copies (e.g., about 5, 10, 20, 35, 50, 75, 100 or 150 copies), of a particular exogenous nucleic acid molecule as described herein.

Methods for expressing an amino acid sequence from an exogenous nucleic acid molecule are well known to those skilled in the art. Such methods include, without limitation, constructing a nucleic acid such that a regulatory element promotes the expression of a nucleic acid sequence that encodes a polypeptide. Typically, regulatory elements are DNA sequences that regulate the expression of other DNA sequences at the level of transcription. Thus, regulatory elements include, without limitation, promoters, enhancers, and the like. Any type of promoter can be used to express an amino acid sequence from an exogenous nucleic acid molecule. Examples of promoters include, without limitation, constitutive promoters, tissue-specific promoters, and promoters responsive or unresponsive to a particular stimulus (e.g., light, oxygen, chemical concentration, and the like). Moreover, methods for expressing a polypeptide from an exogenous nucleic acid molecule in cells such as bacterial cells and yeast cells are well known to those skilled in the art. For example, nucleic acid constructs that are capable of expressing exogenous polypeptides within *E. coli* are well known. See, e.g., Sambrook *et al.*, Molecular cloning: a laboratory manual, Cold Spring Harbour Laboratory Press, New York, USA, second edition (1989).

### B. Production of Organic Acids and Related Products via Host Cells

The nucleic acid and amino acid sequences provided herein can be used with cells to produce 3-HP and/or other organic compounds such as 1,3-propanediol, acrylic acid, polymerized acrylate, esters of acrylate, esters of 3-HP, and polymerized 3-HP. Such cells can be from any species including those listed within the taxonomy web pages at the

National Institute of Health sponsored by the United States government (www.ncbi.nlm.nih.gov). The cells can be eukaryotic or prokaryotic. For example, genetically modified cells of the invention can be mammalian cells (e.g., human, murine, and bovine cells), plant cells (e.g., corn, wheat, rice, and soybean cells), fungal cells (e.g., *Aspergillus* and *Rhizopus* cells), yeast cells, or bacterial cells (e.g., *Lactobacillus*, *Lactococcus*, *Bacillus*, *Escherichia*, and *Clostridium* cells). A cell of the invention also can be a microorganism. The term "microorganism" as used herein refers to any microscopic organism including, without limitation, bacteria, algae, fungi, and protozoa. Thus, *E. coli*, *S. cerevisiae*, *Kluveromyces lactis*, *Candida blankii*, *Candida rugosa*, and *Pichia postoris* are considered microorganisms and can be used as described herein.

Typically, a cell of the invention is genetically modified such that a particular organic compound is produced. In one embodiment, the invention provides cells that make 3-HP from PEP. Examples biosynthetic pathways that cay be used by cells to make 3-HP are shown in Figures 1-5, 43-44, 54, and 55.

Generally, cells that are genetically modified to synthesize a particular organic compound contain one or more exogenous nucleic acid molecules that encode polypeptides having specific enzymatic activities. For example, a microorganism can contain exogenous nucleic acid that encodes a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity. In this case, acrylyl-CoA can be converted into 3-hydroxypropionic acid-CoA which can lead to the production of 3-HP. It is noted that a cell can be given an exogenous nucleic acid molecule that encodes a polypeptide having an enzymatic activity that catalyzes the production of a compound not normally produced by that cell. Alternatively, a cell can be given an exogenous nucleic acid molecule that encodes a polypeptide having an enzymatic activity that catalyzes the production of a compound that is normally produced by that cell. In this case, the genetically modified cell can produce more of the compound, or can produce the compound more efficiently, than a similar cell not having the genetic modification.

In one embodiment, the invention provides a cell containing an exogenous nucleic acid molecule that encodes a polypeptide having enzymatic activity that leads to the formation of 3-HP. It is noted that the produced 3-HP can be secreted from the cell, eliminating the need to disrupt cell membranes to retrieve the organic compound.

Typically, the cell of the invention produces 3-HP with the concentration being at least about 100 mg per L (e.g., at least about 1 g/L, 5 g/L, 10 g/L, 25 g/L, 50 g/L, 75 g/L, 80 g/L, 90 g/L, 100 g/L, or 120 g/L). When determining the yield of an organic compound such as 3-HP for a particular cell, any method can be used. See, e.g., *Applied Environmental Microbiology* 59(12):4261-4265 (1993). Typically, a cell within the scope of the invention such as a microorganism catabolizes a hexose carbon source such as glucose. A cell, however, can catabolize a variety of carbon sources such as pentose sugars (e.g., ribose, arabinose, xylose, and lyxose), fatty acids, acetate, or glycerols. In other words, a cell within the scope of the invention can utilize a variety of carbon sources.

As described herein, a cell within the scope of the invention can contain an exogenous nucleic acid molecule that encodes a polypeptide having enzymatic activity that leads to the formation of 3-HP or other organic compounds such as 1,3-propanediol, acrylic acid, poly-acrylate, acrylate-esters, 3-HP-esters, and poly-3-HP. Methods of identifying cells that contain exogenous nucleic acid are well known to those skilled in the art. Such methods include, without limitation, PCR and nucleic acid hybridization techniques such as Northern and Southern analysis (see hybridization described herein). In some cases, immunohisto-chemistry and biochemical techniques can be used to determine if a cell contains a particular nucleic acid by detecting the expression of the polypeptide encoded by that particular nucleic acid molecule. For example, an antibody having specificity for a polypeptide can be used to determine whether or not a particular cell contains nucleic acid encoding that polypeptide. Further, biochemical techniques can be used to determine if a cell contains a particular nucleic acid molecule encoding a polypeptide having enzymatic activity by detecting an organic product produced as a result of the expression of the polypeptide having enzymatic activity. For example, detection of 3-HP after introduction of exogenous nucleic acid that encodes a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity into a cell that does not normally express such a polypeptide can indicate that that cell not only contains the introduced exogenous nucleic acid molecule but also expresses the encoded polypeptide from that introduced exogenous nucleic acid molecule. Methods for detecting specific enzymatic activities or the presence of particular organic products are well known to those skilled in

the art. For example, the presence of an organic compound such as 3-HP can be determined as described elsewhere. See, Sullivan and Clarke, *J. Assoc. Offic. Agr. Chemists*, 38:514-518 (1955).

### C. Cells with Reduced Polypeptide Activity

The invention also provides genetically modified cells having reduced polypeptide activity. The term "reduced" as used herein with respect to a cell and a particular polypeptide's activity refers to a lower level of activity than that measured in a comparable cell of the same species. For example, a particular microorganism lacking enzymatic activity X is considered to have reduced enzymatic activity X if a comparable microorganism has at least some enzymatic activity X. It is noted that a cell can have the activity of any type of polypeptide reduced including, without limitation, enzymes, transcription factors, transporters, receptors, signal molecules, and the like. For example, a cell can contain an exogenous nucleic acid molecule that disrupts a regulatory and/or coding sequence of a polypeptide having pyruvate decarboxylase activity or alcohol dehydrogenase activity. Disrupting pyruvate decarboxylase and/or alcohol dehydrogenase expression can lead to the accumulation of lactate as well as products produced from lactate such as 3-HP, 1,3-propanediol, acrylic acid, poly-acrylate, acrylate-esters, 3-HP-esters, and poly-3-HP. It is also noted that reduced polypeptide activities can be the result of lower polypeptide concentration, lower specific activity of a polypeptide, or combinations thereof. Many different methods can be used to make a cell having reduced polypeptide activity. For example, a cell can be engineered to have a disrupted regulatory sequence or polypeptide-encoding sequence using common mutagenesis or knock-out technology. See, e.g., Methods in Yeast Genetics (1997 edition), Adams, Gottschling, Kaiser, and Sterns, Cold Spring Harbor Press (1998). Alternatively, antisense technology can be used to reduce the activity of a particular polypeptide. For example, a cell can be engineered to contain a cDNA that encodes an antisense molecule that prevents a polypeptide from being translated. The term "antisense molecule" as used herein encompasses any nucleic acid molecule or nucleic acid analog (e.g., peptide nucleic acids) that contains a sequence that corresponds to the coding strand of an endogenous polypeptide. An antisense molecule also can have

flanking sequences (e.g., regulatory sequences). Thus, antisense molecules can be ribozymes or antisense oligonucleotides. A ribozyme can have any general structure including, without limitation, hairpin, hammerhead, or axhead structures, provided the molecule cleaves RNA. Further, gene silencing can be used to reduce the activity of a particular polypeptide.

A cell having reduced activity of a polypeptide can be identified using any method. For example, enzyme activity assays such as those described herein can be used to identify cells having a reduced enzyme activity.

A polypeptide having (1) the amino acid sequence set forth in SEQ ID NO:39 (the OS17 polypeptide) or (2) an amino acid sequence sharing at least about 60 percent sequence identity with the amino acid sequence set forth in SEQ ID NO:39 can have three functional domains: a domain having CoA-synthatase activity, a domain having 3-HP-CoA dehydratase activity, and a domain having CoA-reductase activity. Such polypeptides can be selectively modified by mutating and/or deleting domains such that one or two of the enzymatic activities are reduced. Reducing the dehydratase activity of the OS17 polypeptide can cause acrylyl-CoA to be created from propionyl-CoA. The acrylyl-CoA then can be contacted with a polypeptide having CoA hydrolase activity to produce acrylate from propionate (Figure 43). Similarly, acrylyl-CoA can be created from 3-HP by using, for example, an OS17 polypeptide having reduced reductase activity.

### D. Production of Organic Acids and Related Products via In Vitro Techniques

In addition, purified polypeptides having enzymatic activity can be used alone or in combination with cells to produce 3-HP or other organic compounds such as 1,3-propanediol, acrylic acid, polymerized acrylate, esters of acrylate, esters of 3-HP, and polymerized 3-HP. For example, a preparation containing a substantially pure polypeptide having 3-hydroxypropionyl-CoA dehydratase activity can be used to catalyze the formation of 3-HP-CoA, a precursor to 3-HP. Further, cell-free extracts containing a polypeptide having enzymatic activity can be used alone or in combination with purified polypeptides and/or cells to produce 3-HP. For example, a cell-free extract containing a

polypeptide having CoA transferase activity can be used to form lactyl-CoA, while a microorganism containing polypeptides have the enzymatic activities necessary to catalyze the reactions needed to form 3-HP from lactyl-CoA can be used to produce 3-HP. Any method can be used to produce a cell-free extract. For example, osmotic shock, sonication, and/or a repeated freeze-thaw cycle followed by filtration and/or centrifugation can be used to produce a cell-free extract from intact cells.

It is noted that a cell, purified polypeptide, and/or cell-free extract can be used to produce 3-HP that is, in turn, treated chemically to produce another compound. For example, a microorganism can be used to produce 3-HP, while a chemical process is used to modify 3-HP into a derivative such as polymerized 3-HP or an ester of 3-HP. Likewise, a chemical process can be used to produce a particular compound that is, in turn, converted into 3-HP or other organic compound (e.g., 1,3-propanediol, acrylic acid, polymerized acrylate, esters of acrylate, esters of 3-HP, and polymerized 3-HP) using a cell, substantially pure polypeptide, and/or cell-free extract described herein. For example, a chemical process can be used to produce acrylyl-CoA, while a microorganism can be used convert acrylyl-CoA into 3-HP.

### E. Fermentation of Cells to Produce Organic Acids

Typically, 3-HP is produced by providing a production cell, such as a microorganism, and culturing the microorganism with culture medium such that 3-HP is produced. In general, the culture media and/or culture conditions can be such that the microorganisms grow to an adequate density and produce 3-HP efficiently. For large-scale production processes, any method can be used such as those described elsewhere (Manual of Industrial Microbiology and Biotechnology, 2nd Edition, Editors: A. L. Demain and J. E. Davies, ASM Press; and Principles of Fermentation Technology, P. F. Stanbury and A. Whitaker, Pergamon). Briefly, a large tank (e.g., a 100 gallon, 200 gallon, 500 gallon, or more tank) containing appropriate culture medium with, for example, a glucose carbon source is inoculated with a particular microorganism. After inoculation, the microorganisms are incubated to allow biomass to be produced. Once a desired biomass is reached, the broth containing the microorganisms can be transferred to a second tank. This second tank can be any size. For example, the second tank can be

larger, smaller, or the same size as the first tank. Typically, the second tank is larger than the first such that additional culture medium can be added to the broth from the first tank. In addition, the culture medium within this second tank can be the same as, or different from, that used in the first tank. For example, the first tank can contain medium with xylose, while the second tank contains medium with glucose.

Once transferred, the microorganisms can be incubated to allow for the production of 3-HP. Once produced, any method can be used to isolate the 3-HP. For example, common separation techniques can be used to remove the biomass from the broth, and common isolation procedures (e.g., extraction, distillation, and ion-exchange procedures) can be used to obtain the 3-HP from the microorganism-free broth. In addition, 3-HP can be isolated while it is being produced, or it can be isolated from the broth after the product production phase has been terminated.

### F. Products Created From the Disclosed Biosynthetic Routes

The organic compounds produced from any of the steps provided in Figures 1-5, 43-44, 54, and 55 can be chemically converted into other organic compounds. For example, 3-HP can be hydrogenated to form 1,3 propanediol, a valuable polyester monomer. Hydrogenating an organic acid such as 3-HP can be performed using any method such as those used to hydrogenate succinic acid and/or lactic acid. For example, 3-HP can be hydrogenated using a metal catalyst. In another example, 3-HP can be dehydrated to form acrylic acid. Any method can be used to perform a dehydration reaction. For example, 3-HP can be heated in the presence of a catalyst (e.g., a metal or mineral acid catalyst) to form acrylic acid. Propanediol also can be created using polypeptides having oxidoreductase activity (e.g., enzymes is the 1.1.1.- class of enzymes) *in vitro* or *in vivo*.

## V. Overview of Methodology Used to Create Biosynthetic Pathways That Make 3-HP from PEP

The invention provides methods of making 3-HP and related products from PEP via the use of biosynthetic pathways. Illustrative examples include methods involving the

production of 3-HP via a lactate intermediate, a malonyl-CoA intermediate, and a ß-alanine intermediate.

### A. Biosynthetic Pathway for Making 3-HP through a Lactic Acid Intermediate

A biosynthetic pathway that allows for the production of 3-HP from PEP was constructed (Figure 1). This pathway involved using several polypeptides that were cloned and expressed as described herein. *M. elsdenii* cells (ATCC 17753) were used as a source of genomic DNA. Primers were used to identify and clone a nucleic acid sequence encoding a polypeptide having CoA transferase activity (SEQ ID NO: 1). The polypeptide was subsequently tested for enzymatic activity and found to have CoA transferase activity.

Similarly, PCR primers were used to identify nucleic acid sequences from *M. elsdenii* genomic DNA that encoded an E1 activator, E2 α, and E2 β polypeptides (SEQ ID NOs: 9, 17, and 25, respectively). These polypeptides were subsequently shown to have lactyl-CoA dehydratase activity.

*Chloroflexus aurantiacus* cells (ATCC 29365) were used as a source of genomic DNA. Initial cloning lead to the identification of nucleic acid sequences: OS17 (SEQ ID NO: 129) and OS19 (SEQ ID NO: 40). Subsequence assays revealed that OS17 encodes a polypeptide having CoA synthase activity, dehydratase activity, and dehydrogenase activity (propionyl-CoA synthatase). Subsequence assays also revealed that OS19 encodes a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity (also referred to as acrylyl-CoA hydratase activity).

Several operons were constructed for use in *E. coli*. These operons allow for the production of 3-HP in bacterial cells. Additional experiments allowed for the expression of these polypeptide is yeast, which can be used to produce 3-HP.

### B. Biosynthetic Pathway for Making 3-HP through a Malonyl-CoA Intermediate

Another pathway leading to the production of 3-HP from PEP was constructed. This pathway used a polypeptide having acetyl CoA carboxylase activity that was isolated

from *E. coli* (Example 9), and a polypeptide having malonyl-CoA reductase activity that was isolated from *Chloroflexus aurantacius* (Example 10). The combination of these two polypeptides allows for the production of 3-HP from acetyl-CoA (Figure 44).

Nucleic acid encoding a polypeptide having malonyl-CoA reductase activity (SEQ ID NO:140) was cloned, sequenced, and expressed. The polypeptide having malonyl-CoA reductase activity was then used to make 3-HP.

### C. Biosynthetic Pathways For Making 3-HP through a ß-alanine Intermediate

In general, prokaryotes and eukaryotes metabolize glucose via the Embden-Meyerhof-Parnas pathway to PEP, a central metabolite in carbon metabolism. The PEP generated from glucose is either carboxylated to oxlaoacetate or is converted to pyruvate. Carboxylation of PEP to oxaloacetate can be catalyzed by a polypeptide having PEP carboxylase activity, a polypeptide having PEP carboxykinase activity, or a polypeptide having PEP transcarboxylase activity. Pyruvate that is generated from PEP by a polypeptide having pyruvate kinase activity can also be converted to oxaloacetate by a polypeptide having pyruvate carboxylase activity.

Oxaloacetate generated either from PEP or pyruvate can act as a precursor for production of aspartic acid. This conversion can be carried out by a polypeptide having aspartate aminotransferase activity, which transfers an amino group from glutamate to oxaloacetate. Glutamate consumed in this reaction can be regenerated by the action of a polypeptide having glutamate dehydrogenase activity or by the action of a polypeptide having 4, 4-aminobutyrate aminotransferase activity. The decarboxylation of aspartate to β-alanine is catalyzed by a polypeptide having aspartate decarboxylase activity. β-alanine produced through this biochemistry can be converted to 3-HP via two possible pathways. These pathways are provided in Figures 54 and 55.

The steps involved in the production of β-alanine can be the same for both pathways. These steps can be accomplished by endogenous polypeptides in the host cells which convert PEP to β-alanine, or these steps can be accomplished with recombinant DNA technology using known polypeptides such as polypeptides having PEP-

carboxykinase activity (4.1.1.32), aspartate aminotransferase activity (2.6.1.1), and aspartate alpha-decarboxylase activity (4.1.1.11).

As depicted in Figure 54, a polypeptide having CoA transferase activity (e.g., a polypeptide having a sequence set forth in SEQ ID NO:2) can be used to convert β-alanine to β-alanyl-CoA. β-alanyl-CoA can be converted to acrylyl-CoA via a polypeptide having β-alanyl-CoA ammonia lyase activity (e.g., a polypeptide having a sequence set forth in SEQ ID NO:160). Acrylyl-CoA can be converted to 3-HP-CoA using a polypeptide having 3-HP-CoA dehydratase activity (e.g., a polypeptide having a sequence set forth in SEQ ID NO:40). 3-HP-CoA can be converted into 3-HP via a polypeptide having CoA transferase activity (e.g., a polypeptide having a sequence set forth in SEQ ID NO:2).

As depicted in Figure 55, a polypeptide having 4,4-aminobutyrate aminotransferase activity (2.6.1.19) can be used to convert β-alanine into malonate semialdehyde. The malonate semialdehyde can be converted into 3-HP using either a polypeptide having 3-hydroxypropionate dehydrogenase activity (1.1.1.59) or a polypeptide having 3-hydroxyisobutyrate dehydrogenase activity.

## EXAMPLES

### Example 1 – Cloning nucleic acid molecules that encode a polypeptide having CoA transferase activity

Genomic DNA was isolated from *Megasphaera elsdenii* cells (ATCC 17753) grown in 1053 Reinforced Clostridium media under anaerobic conditions at 37°C in roll tubes for 12-14 hours. Once grown, the cells were pelleted, washed with 5 mL of a 10 mM Tris solution, and repelleted. The pellet was resuspended in 1 mL of Gentra Cell Suspension Solution to which 14.2 mg of lysozyme and 4 μL of 20 mg/mL proteinase K solution was added. The cell suspension was incubated at 37°C for 30 minutes. The genomic DNA was than isolated using a Gentra Genomic DNA Isolation Kit following the provided protocol. The precipitated genomic DNA was spooled and air-dried for 10 minutes. The genomic DNA was suspended in 500 μL of a 10 mM Tris solution and stored at 4°C.

Two degenerate forward (CoAF1 and CoAF2) and three degenerate reverse (CoAR1, CoAR2, and CoAR3) PCR primers were designed based on conserved acetoacetyl CoA transferase and propionate CoA transferase sequences (CoAF1 5'-GAAWSCGGYSCNATYGGYGG-3', SEQ ID NO: 49; CoAF2 5'-TTYTGYG-GYRSBTTYACBGCWGG-3', SEQ ID NO: 50; CoAR1 5'-CCWGCVGTRAAV-SYRCCRCARAA-3', SEQ ID NO: 51; CoAR2 5'-AARACDSMRCGTTCVGTRA-TRTA-3', SEQ ID NO: 52; and CoAR3 5'-TCRAYRCCSGGWGCRAYTTC-3', SEQ ID NO: 53). The primers were used in all logical combinations in PCR using Taq polymerase (Roche Molecular Biochemicals, Indianapolis, IN) and 1 ng of genomic DNA per μL reaction mix. PCR was conducted using a touchdown PCR program with 4 cycles at an annealing temperature of 59°C, 4 cycles at 57°C, 4 cycles at 55°C, and 18 cycles at 52°C. Each cycle used an initial 30-second denaturing step at 94°C and a 3 minute extension at 72°C. The program had an initial denaturing step for 2 minutes at 94°C and a final extension step of 4 minutes at 72°C. Time allowed for annealing was 45 seconds. The amounts of PCR primer used in the reactions were increased 2-8 fold above typical PCR amounts depending on the amount of degeneracy in the 3' end of the primer. In addition, separate PCR reactions containing each individual primer were made to identify PCR products resulting from single degenerate primers. Each PCR product (25 μL) was separated by electrophoresis using a 1% TAE (Tris-acetate-EDTA) agarose gel.

The CoAF1-CoAR2, CoAF1-CoAR3, CoAF2-CoAR2, and CoAF2-CoAR3 combinations produced a band of 423, 474, 177, and 228 bp, respectively. These bands matched the sizes based on other CoA transferase sequences. No band was visible from the individual primer control reactions. The CoAF1-CoAR3 fragment (474 bp) was isolated and purified using a Qiagen Gel Extraction Kit (Qiagen Inc., Valencia, CA). Four μL of the purified band was ligated into pCRII vector and transformed into TOP10 *E. coli* cells by heat-shock using a TOPO cloning procedure (Invitrogen, Carlsbad, CA). Transformations were plated on LB media containing 100 μg/mL of ampicillin (Amp) and 50 μg/mL of 5-Bromo-4-Chloro-3-Indolyl-*B*-D-Galactopyranoside (X-gal). Single, white colonies were plated onto fresh media and screened in a PCR reaction using the CoAF1 and CoAR3 primers to confirm the presence of the insert.

Plasmid DNA obtained using a QiaPrep Spin Miniprep Kit (Qiagen, Inc) was quantified and used for DNA sequencing with M13R and M13F primers. Sequence analysis revealed that the CoAF1-CoAR3 fragment shared sequence similarity with acetoacetyl CoA transferase sequences.

Genome walking was performed to obtain the complete coding sequence. The following primers for genome walking in both upstream and downstream directions were designed using the portion of the 474 bp CoAF1-CoAR3 fragment sequence that was internal to the degenerate primers (COAGSP1F 5'-GAATGTTTACTTCTGCGG-CACCTTCAC-3', SEQ ID NO:54; COAGSP2F 5'-GACCAGATCACTTTCAACG-GTTCCTATG-3', SEQ ID NO:55; COAGSP1R 5'-GCATAGGAACCGTTGAAA-GTGATCTGG-3', SEQ ID NO:56; and COAGSP2R 5'-GTTAGTACCGAACTTG-CTGACGTTGATG-3', SEQ ID NO:57). The COAGSP1F and COAGSP2F primers face downstream, while the COAGSP1R and COAGSP2R primers face upstream. In addition, the COAGSP2F and COAGSP2R primers are nested inside the COAGSP1F and COAGSP1R primers. Genome walking was performed using the Universal Genome Walking kit (ClonTech Laboratories, Inc., Palo Alto, CA) with the exception that additional libraries were generated with enzymes *Nru* I, *Sca* I, and *Hinc* II. First round PCR was conducted in a Perkin Elmer 2400 Thermocycler with 7 cycles of 2 seconds at 94°C and 3 minutes at 72°C, and 36 cycles of 2 seconds at 94°C and 3 minutes at 65°C with a final extension at 65°C for 4 minutes. Second round PCR used 5 cycles of 2 seconds at 94°C and 3 minutes at 72°C, and 20 cycles of 2 seconds at 94°C and 3 minutes at 65°C with a final extension at 65°C for 4 minutes. The first and second round product (20 μL) was separated by electrophoresis on a 1% TAE agarose gel. Amplification products were obtained with the *Stu* I library for the reverse direction. The second round product of 1.5 Kb from this library was gel purified, cloned, and sequenced. Sequence analysis revealed that the sequence derived from genome walking overlapped with the CoAF1-CoAR3 fragment and shared sequence similarity with other sequences such as acetoacetyl CoA transferase sequences (Figures 8-9).

Nucleic acid encoding the CoA transferase (propionyl-CoA transferase or *pct*) from *Megasphaera elsdenii* was PCR amplified from chromosomal DNA using following PCR program: 25 cycles of 95°C for 30 seconds to denature, 50°C for 30 seconds to

anneal, and 72°C for 3 minutes for extension (plus 2 seconds per cycle). The primers used were designated PCT-1.114 (5'-ATGAGAAAAGTAGAAATCATTAC-3'; SEQ ID NO:58) and PCT-2.2045 (5'-GGCGGAAGTTGACGATAATG-3'; SEQ ID NO:59). The resulting PCR product (about 2 kb as judged by agarose gel electrophoresis) was purified using a Qiagen PCR purification kit (Qiagen Inc., Valencia, CA). The purified product was ligated to pETBlue-1 using the Perfectly Blunt cloning Kit (Novagen, Madison, WI). The ligation reaction was transformed into NovaBlue chemically competent cells (Novagen, Madison, WI) that were spread on LB agar plates supplemented with 50 μg/mL carbenicillin, 40 μg/mL IPTG, and 40 μg/mL X-Gal. White colonies were isolated and screened for the presence of inserts by restriction mapping. Isolates with the correct restriction pattern were sequenced from each end using the primers pETBlueUP and pETBlueDOWN (Novagen) to confirm the sequence at the ligation points.

The plasmid was transformed into Tuner (DE3) pLacI chemically competent cells (Novagen, Madison, WI), and expression from the construct tested. Briefly, a culture was grown overnight to saturation and diluted 1:20 the following morning in fresh LB medium with the appropriate antibiotics. The culture was grown at 37°C with aeration to an $OD_{600}$ of about 0.6. The culture was induced with IPTG at a final concentration of 100 μM. The culture was incubated for an additional two hours at 37°C with aeration. Aliquots were taken pre-induction and 2 hours post-induction for SDS-PAGE analysis. A band of the expected molecular weight (55,653 Daltons predicted from the sequence) was observed after IPTG treatment. This band was not observed in cells containing a plasmid lacking the nucleic acid encoding the transferase.

Cell free extracts were prepared to assess enzymatic activity. Briefly, the cells were harvested by centrifugation and disrupted by sonication. The sonicated cell suspension was centrifuged to remove cell debris, and the supernatant was used in the assays.

Transferase activity was measured in the following assay. The assay mixture used contained 100 mM potassium phosphate buffer (pH 7.0), 200 mM sodium acetate, 1 mM dithiobisnitrobenzoate (DTNB), 500 μM oxaloacetate, 25 μM CoA-ester substrate, and 3 μg/mL citrate synthase. If present, the CoA transferase transfers the CoA from the CoA ester to acetate to form acetyl-CoA. The added citrate synthase condenses oxaloacetate

and acetyl-CoA to form citrate and free CoASH. The free CoASH complexes with DTNB, and the formation of this complex can be measured by a change in the optical density at 412 nm. The activity of the CoA transferase was measured using the following substrates: lactyl-CoA, propionyl-CoA, acrylyl-CoA, and 3-hydroxypropionyl-CoA. The units/mg of protein was calculated using the following formula:

$$(\Delta E/\text{min} * V_f * \text{dilution factor})/ (V_S * 14.2) = \text{units/mL}$$

where $\Delta E/\text{min}$ is the change in absorbance per minute at 412 nm, $V_f$ is the final volume of the reaction, and $V_S$ is the volume of sample added. The total protein concentration of the cell free extract was about 1 mg/mL so the units/mL equals units/mg.

Cell free extracts from cells containing nucleic acid encoding the CoA transferase exhibited CoA transferase activity (Table 2). The observed CoA transferase activity was detected for the lactyl-CoA, propionyl-CoA, acrylyl-CoA, and 3-hydroxypropionyl-CoA substrates (Table 2). The highest CoA transferase activity was detected for lactyl-CoA and propionyl-CoA.

Table 2

| Substrate | Units/mg |
|---|---|
| Lactyl-CoA | 211 |
| Propionyl-CoA | 144 |
| Acrylyl-CoA | 118 |
| 3-Hydroxypropionyl-CoA | 110 |

The following assay was performed to test whether the CoA transferase activity can use the same CoA substrate donors as recipients. Specifically, CoA transferase activity was assessed using a Matrix-assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry (MALDI-TOF MS) Voyager RP workstation (PerSeptive Biosystems). The following five reactions were analyzed:

1) acetate + lactyl-CoA → lactate + acetyl-CoA

2) acetate + propionyl-CoA → propionate + acetyl-CoA

3) lactate + acetyl-CoA → acetate + lactyl-CoA

4) lactate + acrylyl-CoA → acrylate + lactyl-CoA

5) 3-hydroxypropionate + lactyl-CoA → lactate + 3-hydroxypropionyl-CoA

MALDI-TOF MS was used to measure simultaneously the appearance of the product CoA ester and the disappearance of the donor CoA ester. The assay buffer contained 50 mM potassium phosphate (pH 7.0), 1 mM CoA ester, and 100 mM respective acid salt. Protein from a cell free extract prepared as described above was added to a final concentration of 0.005 mg/mL. A control reaction was prepared from a cell free extract prepared from cells lacking the construct containing the CoA transferase-encoding nucleic acid. For each reaction, the cell free extract was added last to start the reaction. Reactions were allowed to proceed at room temperature and were stopped by adding 1 volume 10% trifluroacetic acid (TFA). The reaction mixtures were purified prior to MALDI-TOF MS analysis using Sep Pak Vac $C_{18}$ 50 mg columns (Waters, Inc.). The columns were conditioned with 1 mL methanol and equilibrated with two washes of 1 mL 0.1% TFA. Each sample was applied to the column, and the flow through was discarded. The column was washed twice with 1 mL 0.1% TFA. The sample was eluted in 200 µL 40% acetonitrile, 0.1% TFA. The acetonitrile was removed by centrifugation *in vacuo*. Samples were prepared for MALDI-TOF MS analysis by mixing 1:1 with 110 mM sinapinic acid in 0.1% TFA, 67% acetonitrile. The samples were allowed to air dry.

In reaction #1, the control sample exhibited a main peak at a molecular weight corresponding to lactyl-CoA (MW 841). There was a minor peak at the molecular weight corresponding to acetyl-CoA (MW 811). This minor peak was determined to be the left-over acetyl-CoA from the synthesis of lactyl-CoA. The reaction #1 sample containing the cell extract from cells transfected with the CoA transferase-encoding plasmid exhibited complete conversion of lactyl-CoA to acetyl-CoA. No peak was observed for lactyl-CoA. This result indicates that the CoA transferase activity can transfer CoA from lactyl-CoA to acetate to form acetyl-CoA.

In reaction #2, the control sample exhibited a dominant peak at a molecular weight corresponding to propionyl-CoA (MW 825). The reaction #2 sample containing the cell extract from cells transfected with the CoA transferase-encoding plasmid exhibited a dominant peak at a molecular weight corresponding to acetyl-CoA (MW 811).

No peak was observed for propionyl-CoA. This result indicates that the CoA transferase activity can transfer CoA from propionyl-CoA to acetate to form acetyl-CoA.

In reaction #3, the control sample exhibited a dominant peak at a molecular weight corresponding to acetyl-CoA (MW 811). The reaction #3 sample containing the cell extract from cells transfected with the CoA transferase-encoding plasmid exhibited a peak corresponding to lactyl-CoA (MW 841). The peak corresponding to acetyl-CoA did not disappear. In fact, the ratio of the size of the two peaks was about 1:1. The observed appearance of the peak corresponding to lactyl-CoA demonstrates that the CoA transferase activity catalyzes reaction #3.

In reaction #4, the control sample exhibited a dominant peak at a molecular weight corresponding to acrylyl-CoA (MW 823). The reaction #4 sample containing the cell extract from cells transfected with the CoA transferase-encoding plasmid exhibited a dominant peak corresponding to lactyl-CoA (MW 841). This result demonstrates that the CoA transferase activity catalyzes reaction #4.

In reaction #5, deuterated lactyl-CoA was used to detect the transfer of CoA from lactate to 3-hydroxypropionate since lactic acid and 3-HP have the same molecular weight as do their respective CoA esters. Using deuterated lactyl-CoA allowed for the differentiation between lactyl-CoA and 3-hydroxypropionate using MALDI-TOF MS. The control sample exhibited a diffuse group of peaks at molecular weights ranging from MW 841 to 845 due to the varying amounts of hydrogen atoms that were replaced with deuterium atoms. In addition, a significant peak was observed at a molecular weight corresponding to acetyl-CoA (MW 811). This peak was determined to be the left-over acetyl-CoA from the synthesis of lactyl-CoA. The reaction #5 sample containing the cell extract from cells transfected with the CoA transferase-encoding plasmid exhibited a dominant peak at a molecular weight corresponding to 3-hydroxypropionyl-CoA (MW 841) as opposed to a group of peaks ranging from MW 841 to 845. This result demonstrates that the CoA transferase catalyzes reaction #5.

**Example 2 – Cloning nucleic acid molecules that encode a multiple polypeptide complex having lactyl-CoA dehydratase activity**

The following methods were used to clone an E1 activator polypeptide. Briefly, four degenerate forward and five degenerate reverse PCR primers were designed based on conserved sequences of E1 activator protein homologs (E1F1 5'- GCWACBGGY-TAYGGYCG-3', SEQ ID NO:60; E1F2 5'-GTYRTYGAYRTYGGYGGYCAGGA-3', SEQ ID NO:61; E1F3 5'-ATGAACGAYAARTGYGCWGCWGG-3', SEQ ID NO:62; E1F4 5'-TGYGCWGCWGGYACBGGYCGYTT-3', SEQ ID NO:63; E1R1 5'-TCCT-GRCCRCCRAYRTCRAYRAC-3', SEQ ID NO:64; E1R2 5'-CCWGCWGCRCAY-TTRTCGTTCAT-3', SEQ ID NO:65; E1R3 5'-AARCGRCCVGTRCCWGCWG-CRCA-3', SEQ ID NO:66; E1R4 5'- GCTTCGSWTTCRACRATGSW-3', SEQ ID NO:67; and E1R5 5'-GSWRATRACTTCGCWTTCWGCRAA-3', SEQ ID NO:68).

The primers were used in all logical combinations in PCR using Taq polymerase (Roche Molecular Biochemicals, Indianapolis, IN) and 1 ng of genomic DNA per μL reaction mix. PCR was conducted using a touchdown PCR program with 4 cycles at an annealing temperature of 60°C, 4 cycles at 58°C, 4 cycles at 56°C, and 18 cycles at 54°C. Each cycle used an initial 30-second denaturing step at 94°C and a 3 minute extension step at 72°C. The program had an initial denaturing step for 2 minutes at 94°C and a final extension step of 4 minutes at 72°C. Time allowed for annealing was 45 seconds. The amounts of PCR primer used in the reactions were increased 2-10 fold above typical PCR amounts depending on the amount of degeneracy in the 3' end of the primer. In addition, separate PCR reactions containing each individual primer were made to identify PCR product resulting from single degenerate primers. Each PCR product (25 μL) was separated by electrophoresis using a 1% TAE (Tris-acetate-EDTA) agarose gel.

The E1F2-E1R4, E1F2-E1R5, E1F3-E1R4, E1F3-E1R5, and E1F4-E1R4R2 combinations produced a band of 195, 207, 144, 156, and 144 bp, respectively. These bands matched the expected size based on E1 activator sequences from other species. No band was visible with individual primer control reactions. The E1F2-E1R5 fragment (207 bp) was isolated and purified using Qiagen Gel Extraction procedure (Qiagen Inc., Valencia, CA). The purified band (4 μL) was ligated into a *pCRII* vector that then was transformed into TOP10 *E. coli* cells by heat-shock using a TOPO cloning procedure

(Invitrogen, Carlsbad, CA). Transformations were plated on LB media containing 100 μg/mL of ampicillin (Amp) and 50 μg/mL of 5-Bromo-4-Chloro-3-Indolyl-*B*-D-Galactopyranoside (X-gal). Single, white colonies were plated onto fresh media and screened in a PCR reaction using the E1F2 and E1R5 primers to confirm the presence of the insert. Plasmid DNA was obtained from multiple colonies using a QiaPrep Spin Miniprep Kit (Qiagen, Inc). Once obtained, the plasmid DNA was quantified and used for DNA sequencing with M13R and M13F primers. Sequence analysis revealed a nucleic acid sequence encoding a polypeptide and revealed that the E1F2-E1R5 fragment shared sequence similarity with E1 activator sequences (Figures 12-13).

Genome walking was performed to obtain the complete coding sequence of E2 α and β subunits. Briefly, four primers for performing genome walking in both upstream and downstream directions were designed using the portion of the 207 bp E1F2-E1R5 fragment sequence that was internal to the E1F2 and E1R5 degenerate primers (E1GSP1F 5'-ACGTCATGTCGAAGGTACTGGAAATCC-3', SEQ ID NO:69; E1GSP2F 5'-GGGACTGGTACTTCAAATCGAAGCATC-3', SEQ ID NO:70; E1GSP1R 5'-TGACGGCAGCGGGATGCTTCGATTTGA-3', SEQ ID NO:71; and E1GSP2R 5'-TCAGACATGGGGATTTCCAGTACCTTC-3', SEQ ID NO:72). The E1GSP1F and E1GSP2F primers face downstream, while the E1GSP1R and E1GSP2R primers face upstream. In addition, the E1GSP2F and E1GSP2R primers are nested inside the E1GSP1F and E1GSP1R primers.

Genome walking was performed using the Universal Genome Walking Kit (ClonTech Laboratories, Inc., Palo Alto, CA) with the exception that additional libraries were generated with enzymes *Nru* I, *Sca* I, and *Hinc* II. First round PCR was performed in a Perkin Elmer 2400 Thermocycler with 7 cycles of 2 seconds at 94°C and 3 minutes at 72°C, and 36 cycles of 2 seconds at 94°C and 3 minutes at 65°C with a final extension at 65°C for 4 minutes. Second round PCR used 5 cycles of 2 seconds at 94°C and 3 minutes at 72°C, and 20 cycles of 2 seconds at 94°C and 3 minutes at 65°C with a final extension at 65°C for 4 minutes. The first and second round product (20 μL) was separated by electrophoresis using 1% TAE agarose gel. Amplification products were obtained with the *Stu* I library for both forward and reverse directions. The second round product of about 1.5 kb for forward direction and 3 kb fragment for reverse direction from the *Stu* I

library were gel purified, cloned, and sequenced. Sequence analysis revealed that the sequence derived from genome walking overlapped with the E1F2-E1R5 fragment.

To obtain additional sequence, a second genome walk was performed using a first round primer (E1GSPF5 5'-CCGTGTTACTTGGGAAGGTATCGCTGTCTG-3', SEQ ID NO:73) and a second round primer (E1GSPF6 5'-GCCAATGAAGGAGGAAA-CCACTAATGAGTC-3', SEQ ID NO:74). The genome walk was performed using the *Nru*I, *Sca*I, and *Hinc*II libraries. In addition, ClonTech's Advantage-Genomic Polymerase was used for the PCR. First round PCR was performed in a Perkin Elmer 2400 Thermocycler with an initial denaturing step at 94°C for 2 minutes, 7 cycles of 2 seconds at 94°C and 3 minutes at 72°C, and 36 cycles of 2 seconds at 94°C and 3 minutes at 65°C with a final extension at 65°C for 4 minutes. Second round PCR used 5 cycles of 2 seconds at 94°C and 3 minutes at 72°C, and 20 cycles of 2 seconds at 94°C and 3 minutes at 65°C with a final extension at 65°C for 4 minutes. The first and second round product (20 μL) was separated by electrophoresis on a 1% agarose gel. An about 1.5 kb amplification product was obtained from second round PCR of the *Hinc*II library. This band was gel purified, cloned, and sequenced. Sequence analysis revealed that it overlapped with the previously obtained genome walk fragment. In addition, sequence analysis revealed a nucleic acid sequence encoding an E2 α subunit that shares sequence similarities with other sequences (Figures 16-17). Further, sequence analysis revealed a nucleic acid sequence encoding an E2 β subunit that shares sequence similarities with other sequences (Figures 20-21).

Additional PCR and sequence analysis revealed the order of polypeptide encoding sequences within the region containing the lactyl-CoA dehydratase-encoding sequences. Specifically, the E1GSP1F and COAGSP1R primer pair and the COAGSP1F and E1GSP1R primer pair were used to amplify fragments that encode both the CoA transferase and E1 activator polypeptides. Briefly, *M. elsdenii* genome DNA (1 ng) was used as a template. The PCR was conducted in Perkin Elmer 2400 Thermocycler using Long Template Polymerase (Roche Molecular Biochemicals, Indianapolis, IN). The PCR program used was as follows: 94°C for 2 minutes; 29 cycles of 94°C for 30 seconds, 61°C for 45 seconds, and 72°C for 6 minutes; and a final extension of 72°C for 10 minutes. Both PCR products (20 μL) were separated on a 1% agarose gel. An

amplification product (about 1.5 kb) was obtained using the COAGSP1F and E1GSP1R primer pair. This product was gel purified, cloned, and sequenced (Figure 22).

The organization of the *M. elsdenii* operon containing the lactyl-CoA dehydratase-encoding sequences was determined to containing the following polypeptide-encoding sequences in the following order: CoA transferase (Figure 6), ORFX (Figure 23), E1 activator protein of lactyl-CoA dehydratase (Figure 10), E2 α subunit of lactyl-CoA dehydratase (Figure 14), E2 β subunit of lactyl-CoA dehydratase (Figure 18), and truncated CoA dehydrogenase (Figure 25).

The lactyl-CoA dehydratase (lactyl-CoA dehydratase or *lcd*) from *M. elsdenii* was PCR amplified from chromosomal DNA using the following program: 94°C for 2 minutes; 7 cycles of 94°C for 30 seconds, 47°C for 45 seconds, and 72°C for 3 minutes; 25 cycles of 94°C for 30 seconds, 54°C for 45 seconds, and 72°C for 3 minutes; and 72°C for 7 minutes. One primer pair was used (OSNBE1F 5'-GGGAATTCCATATG-AAAACTGTGTATACTCTC-3', SEQ ID NO:75 and OSNBE1R 5'-CGACGGAT-CCTTAGAGGATTTCCGAGAAAGC-3', SEQ ID NO:76). The amplified product (about 3.2 kb) was separated on 1% agarose gel, cut from the gel, and purified with a Qiagen Gel Extraction kit (Qiagen, Valencia, CA). The purified product was digested with *Nde* I and *Bam*HI restriction enzymes and ligated into pET11a vector (Novagen) digested with the same enzymes. The ligation reaction was transformed into NovaBlue chemically competent cells (Novagen) that then were spread on LB agar plates supplemented with 50 μg/mL carbenicillin. Isolated individual colonies were screened for the presence of inserts by restriction mapping. Isolates with the correct restriction pattern were sequenced from each end using Novagen primers (T7 promoter primer #69348-3 and T7 terminator primer #69337-3) to confirm the sequence at the ligation points.

A plasmid having the correct insert was transformed into Tuner (DE3) pLacI chemically competent cells (Novagen, Madison, WI). Expression from this construct was tested as follows. A culture was grown overnight to saturation and diluted 1:20 the following morning in fresh LB medium with the appropriate antibiotics. The culture was grown at 37°C with aeration to an $OD_{600}$ of about 0.6. The culture was induced with IPTG at a final concentration of 100 μM. The culture was incubated for an additional two

hours at 37°C with aeration. Aliquots were taken pre-induction and 2 hours post-induction for SDS-PAGE analysis. Bands of the expected molecular weight (27,024 Daltons for the E1 subunit, 48,088 Daltons for the E2 α subunit, and 42,517 Daltons for the E2 β subunit—all predicted from the sequence) were observed. These bands were not observed in cells containing a plasmid lacking the nucleic acid encoding the three components of the lactyl-CoA dehydratase.

Cell free extracts were prepared by growing cells in a sealed serum bottle overnight at 37°C. Following overnight growth, the cultures were induced with 1 mM IPTG (added using anaerobic technique) and incubated an additional 2 hours at 37°C. The cells were harvested by centrifugation and disrupted by sonication under strict anaerobic conditions. The sonicated cell suspension was centrifuged to remove cell debris, and the supernatant was used in the assays. The buffer used for cell resuspension/sonication was 50 mM Tris-HCl (pH 7.5), 200 μM ATP, 7 mM $Mg(SO_4)$, 4 mM DTT, 1 mM dithionite, and 100 μM NADH.

Dehydratase activity was detected with MALDI-TOF MS. The assay was conducted in the same buffer as above with 1 mM lactyl-CoA or 1 mM acrylyl-CoA added and about 5 mg/mL cell free extract. Prior to MALDI-TOF MS analysis, samples were purified using Sep Pak Vac $C_{18}$ columns (Waters, Inc.) as described in Example 1. The following two reactions were analyzed:

1) acrylyl-CoA → lactyl-CoA

2) lactyl-CoA → acrylyl-CoA

In reaction #1, the control sample exhibited a peak at a molecular weight corresponding to acrylyl-CoA (MW 823). The reaction #1 sample containing the cell extract from cells transfected with the dehydratase-encoding plasmid exhibited a major peak at a molecular weight corresponding to lactyl-CoA (MW 841). This result indicates that the dehydratase activity can convert acrylyl-CoA into lactyl-CoA.

To detect dehydratase activity on lactyl-CoA, reaction #2 was carried out in 80% $D_2O$. The control sample exhibited a peak at a molecular weight corresponding to lactyl-CoA (MW 841). The reaction #2 sample containing the cell extract from cells transfected with the dehydratase-encoding plasmid revealed a lactyl-CoA peak shifted to a deuterated

form. This result indicates that the dehydratase enzyme is active on lactyl-CoA. In addition, the results from both reactions indicate that the dehydratase enzyme can catalyze the lactyl-CoA ←→ acrylyl-CoA reaction in both directions.

**Example 3 – Cloning nucleic acid molecules that encode a polypeptide having 3-hydroxypropionyl CoA dehydratase activity**

Genomic DNA was isolated from *Chloroflexus aurantiacus* cells (ATCC 29365). Briefly, *C. aurantiacus* cells in 920 Chloroflexus medium were grown in 50 mL cultures (Falcon 2070 polypropylene tubes) using an Innova 4230 Incubator, Shaker (New Brunswick Scientific; Edison, NJ) at 50°C with interior lights. Once grown, the cells were pelleted, washed with 5 mL of a 10 mM Tris solution, and re-pelleted. Genomic DNA was isolated from the pelleted cells using a Gentra Genomic "Puregene" DNA isolation kit (Gentra Systems; Minneapolis, MN). Briefly, the pelleted cells were resuspended in 1 mL Gentra Cell Suspension Solution to which 14.2 mg of lysozyme and 4 μL of 20 mg/mL proteinase K solution was added. The cell suspension was incubated at 37°C for 30 minutes. The precipitated genomic DNA was recovered by centrifugation at 3500 x *g* for 25 minutes and air-dried for 10 minutes. The genomic DNA was suspended in 300 μL of a 10 mM Tris solution and stored at 4°C.

The genomic DNA was used as a template in PCR amplification reactions with primers designed based on conserved domains of crotonase homologs and a *Chloroflexus aurantiacus* codon usage table. Briefly, two degenerate forward (CRF1 and CRF2) and three degenerate reverse (CRR1, CRR2, and CRR3) PCR primers were designed (CRF1 5'-AAYCGBCCVAARGCNCTSAAYGC-3', SEQ ID NO:77; CRF2: 5'-TTYGTBGCNGGYGCNGAYAT-3', SEQ ID NO:78; CRR1 5'-ATRTCNG-CRCCNGCVACRAA-3', SEQ ID NO:79; CRR2 5'-CCRCCRCCSAGNG-CRWARCCRTT-3', SEQ ID NO:80; and CRR3 5'-SSWNGCRATVCGRATRTCRAC-3', SEQ ID NO:81).

These primers were used in all logical combinations in PCR using Taq polymerase (Roche Molecular Biochemicals; Indianapolis, IN) and 1 ng of the genomic DNA per μL reaction mix. The PCR was conducted using a touchdown PCR program with 4 cycles at an annealing temperature of 61°C, 4 cycles at 59°C, 4 cycles at 57°C, 4 cycles at 55°C,

and 16 cycles at 52°C. Each cycle used an initial 30-second denaturing step at 94°C and a 3-minute extension step at 72°C. The program also had an initial denaturing step for 2 minutes at 94°C and a final extension step of 4 minutes at 72°C. The time allowed for annealing was 45 seconds. The amounts of PCR primer used in the reaction were increased 4-12 fold above typical PCR amounts depending on the amount of degeneracy in the 3' end of the primer. In addition, separate PCR reactions containing each individual primer were performed to identify amplification products resulting from single degenerate primers. Each PCR product (25 μL) was separated by gel electrophoresis using a 1% TAE (Tris-acetate-EDTA) agarose gel.

The CRF1-CRR1 and CRF2-CRR2 combinations produced a unique band of about 120 and about 150 bp, respectively. These bands matched the expected size based on crotonase genes from other species. No 120 bp or 150 bp band was observed from individual primer control reactions. Both fragments (i.e., the 120 bp and 150 bp bands) were isolated and purified using the Qiagen Gel Extraction kit (Qiagen Inc., Valencia, CA). Each purified fragment (4 μL) was ligated into *pCRII* vector that then was transformed into TOP10 *E. coli* cells by a heat-shock method using a TOPO cloning procedure (Invitrogen, Carlsbad, CA). Transformations were plated on LB media containing 100 μg/mL of ampicillin (Amp) and 50 μg/mL of 5-Bromo-4-Chloro-3-Indolyl-*B*-D-Galactopyranoside (X-gal). Single, white colonies were plated onto fresh media and screened in a PCR reaction using the CRF1 and CRR1 primers and the CRF2 and CRR2 primers to confirm the presence of the desired insert. Plasmid DNA was obtained from multiple colonies with the desired insert using a QiaPrep Spin Miniprep Kit (Qiagen, Inc.). Once obtained, the DNA was quantified and used for DNA sequencing with M13R and M13F primers. Sequence analysis revealed the presence of two different clones from the PCR product of about 150 bp. Each shared sequence similarity with crotonase and hydratase sequences. The two clones were designated OS17 (157 bp PCR product) and OS19 (151 bp PCR product).

Genome walking was performed to obtain the complete coding sequence of OS17. Briefly, primers for conducting genome walking in both upstream and downstream directions were designed using the portion of the 157 bp CRF2-CRR2 fragment sequence that was internal to the CRF2 and CRR2 degenerate primers (OS17F1 5'-CGCTG-

ATATTCGCCAGTTGCTCGAAG-3', SEQ ID NO:82; OS17F2 5'-CCCATCTTG-CTTTCCGCAAGATTGAGC-3', SEQ ID NO:83; OS17F3 5'-CAATGGCCCTGCCGA-ATAACGCCCATCT-3', SEQ ID NO:84; OS17R1 5'-CTTCGAGCAACTGGCGAA-TATCAGCG-3', SEQ ID NO:85; OS17R2 5'-GCTCAATCTTGCGGAAAGCAAG-ATGGG-3', SEQ ID NO:86; and OS17R3 5'-AGATGGGCGTTATTCGGCAGGGCC-ATTG-3', SEQ ID NO:87). The OS17F1, OS17F3, and OS17F2 primers face downstream, while the OS17R2, OS17R3, and OS17R1 primers face upstream.

Genome walking was conducted using the Universal Genome Walking kit (ClonTech Laboratories, Inc., Palo Alto, CA) with the exception that additional libraries were generated with enzymes *Nru* I, *Fsp* I, and *Hinc* II. The first round PCR was conducted in a Perkin Elmer 2400 Thermocycler with 7 cycles of 2 seconds at 94°C and 3 minutes at 72°C, and 36 cycles of 2 seconds at 94°C and 3 minutes at 66°C with a final extension at 66°C for 4 minutes. Second round PCR used 5 cycles of 2 seconds at 94°C and 3 minutes at 72°C, and 20 cycles of 2 seconds at 94°C and 3 minutes at 66°C with a final extension at 66°C for 4 minutes. The first and second round amplification product (5 μL) was separated by gel electrophoresis on a 1% TAE agarose gel. After the second round PCR, an amplification product of about 0.4 kb was obtained with the *Fsp* I library using the OS17R1 primer in the reverse direction, and an amplification product of about 0.6 kb was obtained with the *Hinc* II library using the OS17F2 primer in the forward direction. These PCR products were cloned and sequenced.

Sequence analysis revealed that the sequences derived from genome walking overlapped with the CRF2-CRR2 fragment and shared sequence similarity with crotonase and hydratase sequences.

A second genome walking was performed to obtain additional sequences. Six primers were designed for this second genome walk (OS17F4 5'-AAGCTGGG-TCTGATCGATGCCATTGCTACC-3', SEQ ID NO:88; OS17F5 5'-CTCGATTATCG-CCCATCCACGTATCGAG-3', SEQ ID NO:89; OS17F6 5'-TGGATGCAATCCG-CTATGGCATTATCCACG-3', SEQ ID NO:90; OS17R4 5'-TCATTCAGTGCG-TTCACCGGCGGATTTGTC-3', SEQ ID NO:91; OS17R5 5'-TCGATCCGGAAGT-AGCGATAGCGTTCGATG-3', SEQ ID NO:92; and OS17R6 5'-CTTGGCTGCAAT-CTCTTCGAGCACTTCAGG-3', SEQ ID NO:93). The OS17F4, OS17F5, and OS17F6

primers faced downstream, while the OS17R4, OS17R5, and OS17R6 primers faced upstream.

The second genome walk was performed using the same methods described for the first genome walk. After the second round of walking, an amplification product of about 2.3 kb was obtained with a *Hinc II* library using the OS17R5 primer in the reverse direction, and an amplification product of about 0.6 kb was obtained with a *Pvu* II library using the OS17F5 primer in the forward direction. The PCR products were cloned and sequenced. Sequence analysis revealed that the sequences derived from the second genome walking overlapped with the sequence obtained during the first genome walking. In addition, the sequence analysis revealed a sequence with 3572 bp.

A BLAST search revealed that the polypeptide encoded by this sequence shares sequence similarity with polypeptides having three different activities. Specifically, the beginning of the OS17 encoded-polypeptide shares sequence similarity with CoA-synthesases, the middle region of the OS17 encoded-polypeptide shares sequence similarity with enoyl-CoA hydratases, and the end region of the OS17 encoded-polypeptide shares sequence similarity with CoA-reductases.

A third genome walk was performed using four primers (OS17UP-6 5'-CATCAGAGGTAATCACCACTCGTGCA-3', SEQ ID NO:94; OS17UP-7 5'-AAGTAGTAGGCCACCTCGTCGCCATA-3', SEQ ID NO:95; OS17DN-1 5'-GCCAATCAGGCGCTGATCTATGTTCT-3', SEQ ID NO:96; and OS17DN-2 5'-CTGATCTATGTTCTGGCCTCGGAGGT-3', SEQ ID NO:97). The OS17UP-6 and OS17UP-7 primers face upstream, while the OS17DN-1 and OS17DN-2 primers face downstream. The third genome walk yielded an amplification product of about 1.2 kb with a *Nru* I library using the OS17UP-7 primer in the reverse direction. In addition, amplification products of about 4 kb and about 1.1 kb were obtained with a *Hinc* II and *Fsp* I library, respectively, using the OS17DN-2 primer in the forward direction. Sequence analysis revealed a nucleic acid sequence encoding a polypeptide (Figures 27-28). The complete OS17 gene had 5466 nucleotides and encoded a 1822 amino acid polypeptide. The calculated molecular weight of the OS17 polypeptide from the sequence was 201,346 (pI=5.71).

A BLAST search analysis revealed that the product of the OS17 nucleic acid has three different activities based on sequence similarity to (1) CoA-synthesases at the beginning of the OS17 sequence, (2) 3-HP dehydratases in the middle of the OS17 sequence, and (3) CoA-reductases at the end of the OS17 sequence. Thus, the OS17 clone appeared to encode a single enzyme capable of catalyzing three distinct reactions leading to the direct conversion of 3-hydroxypropionate to propionyl CoA: 3-HP→ 3-HP-CoA→ acrylyl-CoA→propionyl-CoA.

The OS17 gene from *C. aurantiacus* was PCR amplified from chromosomal DNA using the following conditions: 94°C for 3 minutes; 25 cycles of 94°C for 30 seconds to denature, 54°C for 30 seconds to anneal, and 68°C for 6 minutes for extension; followed by 68°C for 10 minutes for final extension. Two primers were used (OS17F 5'-GGGAATTCCATATGATCGACACTGCG-3', SEQ ID NO:136; and OS17R 5'-CGAAGGATCCAACGATAATCGGCTCAGCAC-3', SEQ ID NO:137). The resulting PCR product (~5.6 Kb) was purified using Qiagen PCR purification kit (Qiagen Inc., Valencia, CA). The purified product was digested with NdeI and BamHI restriction enzymes, heated at 80°C for 20 minutes to inactivate the enzymes; purified using Qiagen PCR purification kit, and ligated into a pET11a vector (Novagen, Madison, WI) previously digested with NdeI and BamHI enzymes. The ligation reaction was transformed into NovaBlue chemically competent cells (Novagen, Madison, WI) that were spread on LB agar plates supplemented with 50 μg/mL carbenicillin. Individual transformants were screened by PCR amplification of the OS17 DNA with the OS17F and OS17R primers and conditions as described above directly from colonies cells. Clones that yielded the 5.6 Kb product were used for plasmid purification with Qiagen QiaPrep Spin Miniprep Kit (Qiagen, Inc). Resulting plasmids were transformed into *E. coli* BL21(DE3) cells, and OS17 polypeptide expression induced. The apparent molecular weight of the OS17 polypeptide according to SDS gel electrophoresis was about 190,000 Da.

To assay OS17 polypeptide function, a 100 mL culture of BL21-DE3/pET11a-OS17 cells was started using 1 mL of overnight grown culture as an inoculum. The culture was grown to an OD of 0.5-0.6 and was induced with 100 μM IPTG. After two and a half hours of induction, the cells were harvested by spinning at 8000 rpm in the

floor centrifuge. The cells were washed with 10 mM Tris-HCl (pH 7.8) and passed twice through a French Press at a gauge pressure of 1000 psi. The cell debris was removed by centrifugation at 15,000 rpm. The activity of the OS17 polypeptide was measured spectrophotometrically, and the products formed during this enzymatic transformation were detected by LC/MS. The assay mix was as follows (*J. Bacteriol.*, 181:1088-1098 (1999)):

| Reagent | Volume | Final Conc. |
|---|---|---|
| Tris-HCl (1000 mM, 7.8 pH) | 10 µL | 50 mM |
| $MgCl_2$ (100mM) | 10 µL | 5 mM |
| ATP (30 mM) | 20 µL | 3 mM |
| KCl (100 mM) | 20 µL | 10 mM |
| CoASH (5 mM) | 20 µL | 0.5 mM |
| NAD(P)H | 20 µL | 0.5 mM |
| 3-hydroxypropionate | 2 µL | 1 mM |
| Protein extract (7 mg/mL) | 20 (40 ) µL | 140 µg |
| DI water | 78 (58) µL | |
| Total | 200 µL | |

The initial rate of reaction was measured by monitoring the disappearance of NAD(P)H at 340 nm. The activity of the OS17 polypeptide was measured using 3-HP as the substrate. The units/mL of total protein was calculated using the formula set forth in Example 1. The activity of the expressed OS17 polypeptide was calculated to be 0.061 U/mL of total protein. The reaction products were purified using a Sep Pak Vac column (Waters). The column was conditioned with 1 mL methanol and washed two times with 0.5 mL 0.1% TFA. The sample was then applied to the column, and the column was washed two more times with 0.5 mL 0.1% TFA. The sample was eluted with 200 µL of 40% acetonitrile, 0.1% TFA. The acetonitrile was removed from the sample by vacuum centrifugation. The reaction products were analyzed by LC/MS.

Analyses of thioesters namely propionyl CoA, acrylyl CoA, and 3 HP CoA from the above reaction were carried out using a Waters/Micromass ZQ LC/MS instrument

which had a Waters 2690 liquid chromatograph with a Waters 996 Photo-Diode Array (PDA) placed in series between the chromatograph and the single quadropole mass spectrometer. LC separations were made using a 4.6 x 150 mm YMC ods-AQ (3 μm particles, 120 Å pores) reversed-phase chromatography column at room temperature. CoA esters were eluted in Buffer A (25 mM ammonium acetate, 0.5% acetic acid) with a linear gradient of buffer B (acetonitrile, 0.5% acetic acid). A flow rate of 0.25 mL/minute was used, and photodiode array UV absorbance was monitored from 200 to 400 nm. All parameters of the electrospray MS system were optimized and selected based on generation of protonated molecular ions ($[M+H]^+$) of the analytes of interest and production of characteristic fragment ions. The following instrumental parameters were used for ESI-MS detection of CoA and organic acid-CoA thioesters in the positive ion mode; Extractor: 1 V; RF lens: 0 V; Source temperature: 100°C; Desolvation temperature: 300°C; Desolvation gas: 500 L/hour; Cone gas: 40 L/hour; Low mass resolution: 13.0; High mass resolution: 14.5; Ion energy: 0.5; Multiplier: 650. Uncertainties for mass charge ratios (*m/z*) and molecular masses are ± 0.01%.

The enzyme assay mix from strains expressing the OS17 polypeptide exhibited peaks for propionyl CoA, acrylyl CoA, and 3-HP CoA with the propionyl CoA peak being the dominant peak. These peaks where missing in the enzyme assay mix obtained from the control strain, which carried vector pET11a without an insert. These results indicate that the OS17 polypeptide has CoA synthetase activity, CoA hydratase activity, and dehydrogenase activity.

Genome walking also was performed to obtain the complete coding sequence of OS19. Briefly, primers for conducting genome walking in both upstream and downstream directions were designed using the portion of the 151 bp CRF2-CRR2 fragment sequence that was internal to the CRF2 and CRR2 degenerate primers (OS19F1 5'-GGCTGATATCAAAGCGATGGCCAATGC-3', SEQ ID NO:98; OS19F2 5'-CCAC-GCCTATTGATATGCTCACCAGTG-3', SEQ ID NO:99; OS19F3 5'-GCAAACCGG-TGATTGCTGCCGTGAATGG-3', SEQ ID NO:100; OS19R1 5'-GCATTGGCCAT-CGCTTTGATATCAGCC-3', SEQ ID NO:101; OS19R2 5'-CACTGGTGAGCATATC-AATAGGCGTGG-3', SEQ ID NO:102; and OS19R3 5'-CCATTCACGGCAGCAA-

TCACCGGTTTGC-3', SEQ ID NO:103). The OS19F1, OS19F2, and OS19F3 primers face downstream, while the OS19R1, OS19R2, and OS19R3 primers face upstream.

An amplification product of about 0.25 kb was obtained with the *Fsp* I library using the OS19R1 primer, while an amplification product of about 0.65 kb was obtained with the *Pvu* II library using the OS19R1 primer. In addition, an amplification product of about 0.4 kb was obtained with the *Pvu* II library using the OS19F3 primer. The PCR products were cloned and sequenced. Sequence analysis revealed that the sequences derived from genome walking overlapped with the CRF2-CRR2 fragment and shared sequence similarity with crotonase and hydratase sequences. The obtained sequences accounted for most of the coding sequence including the start codon.

A second genome walk was performed to obtain additional sequence using two primers (OS19F7 5'-TCATCATCGCCAGTGAAAACGCGCAGTTCG-3', SEQ ID NO:104 and OS19F8 5'-GGATCGCGCAAACCATTGCCACCAAATCAC-3', SEQ ID NO:105). The OS19F7 and OS19F8 primers face downstream.

An amplification product (about 0.7 kb) obtained from the *Pvu* II library was cloned and sequenced. Sequence analysis revealed that the sequence derived from the second genome walk overlapped with the sequence obtained from the first genome walk and contained the stop codon. The full-length OS19 clone was found to share sequence similarity with other sequences such as crotonase and enoyl-CoA hydratase sequences (Figures 32-33).

The OS19 clone was found to encode a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity also referred to as acrylyl-CoA hydratase activity. The nucleic acid encoding the OS19 dehydratase from C. *aurantiacus* was PCR amplified from chromosomal DNA using the following conditions: 94°C for 3 minutes; 25 cycles of 94°C for 30 seconds to denature, 56°C for 30 seconds to anneal, and 68°C for 1 minute for extension; and 68°C for 5 minutes for final extension. Two primers were used (OSACH3 5'-ATGAGTGAAGAGTCTCTGGTTCTCAGC-3', SEQ ID NO:106 and OSACH2 5'-AGATCGCAATCGCTCGTGTATGTC-3', SEQ ID NO:107).

The resulting PCR product (about 1.2 kb) was separated by agarose gel electrophoresis and purified using Qiagen PCR purification kit (Qiagen Inc.; Valencia, CA). The purified product was ligated into pETBlue-1 using the Perfectly Blunt cloning

Kit (Novagen; Madison, WI). The ligation reaction was transformed into NovaBlue chemically competent cells (Novagen, Madison, WI) that then were spread on LB agar plates supplemented with 50 μg/mL carbenicillin, 40 μg/mL IPTG, and 40 μg/mL X-Gal. White colonies were isolated and screened for the presence of inserts by restriction mapping. Isolates with the correct restriction pattern were sequenced from each end using the primer pETBlueUP and pETBlueDOWN (Novagen) to confirm the sequence at the ligation points.

The plasmid containing the OS19 dehydratase-encoding sequence was transformed into Tuner (DE3) pLacI chemically competent cells (Novagen, Madison, WI), and expression from the construct tested. Briefly, a culture was grown overnight to saturation and diluted 1:20 the following morning in fresh LB medium with the appropriate antibiotics. The culture was grown at 37°C and 250 rpm to an $OD_{600}$ of about 0.6. At this point, the culture was induced with IPTG at a final concentration of 1 mM. The culture was incubated for an additional two hours at 37°C and 250 rpm. Aliquots were taken pre-induction and 2 hours post-induction for SDS-PAGE analysis. A band of the expected molecular weight (27,336 Daltons predicted from the sequence) was observed. This band was not observed in cells containing a plasmid lacking the nucleic acid encoding the hydratase.

Cell free extracts were prepared by growing cells as described above. The cells were harvested by centrifugation and disrupted by sonication. The sonicated cell suspension was centrifuged to remove cell debris, and the supernatant was used in the assays. The ability of the 3-hydroxypropionyl-CoA dehydratase to perform the following three reactions was measured using MALDI-TOF MS:

1) acrylyl-CoA → 3-hydroxypropionyl-CoA
2) 3-hydroxypropionyl-CoA → acrylyl-CoA
3) crotonyl-CoA → 3-hydroxybutyryl-CoA

The assay mixture contained 50 mM Tris-HCl (pH 7.5), 1 mM CoA ester, and about 1 μg cell free extract. Reactions were allowed to proceed at room temperature and were stopped by adding 1 volume 10% trifluroacetic acid (TFA). The reaction mixtures were purified prior to MALDI-TOF MS analysis using Sep Pak Vac $C_{18}$ 50 mg columns

(Waters, Inc.). The columns were conditioned with 1 mL methanol and then equilibrated with two washes of 1 mL 0.1% TFA. The sample was applied to the column, and the flow through was discarded. The column was washed twice with 1 mL 0.1% TFA. The sample was eluted in 200 µL 40% acetonitrile, 0.1% TFA. The acetonitrile was removed by centrifugation *in vacuo*. Samples were prepared for MALDI-TOF MS analysis by mixing 1:1 with 110 mM sinapinic acid in 0.1% TFA, 67% acetonitrile. The samples were allowed to air dry.

The conversion of acrylyl-CoA into 3-hydroxypropionyl-CoA catalyzed by the 3-hydroxypropionyl-CoA dehydratase was detected using the MALDI-TOF MS technique. In reaction #1, the control sample exhibited a dominant peak at a molecular weight corresponding to acrylyl-CoA (MW 823). The reaction #1 sample containing the cell extract from cells transfected with the 3-hydroxypropionyl-CoA dehydratase-encoding plasmid exhibited a dominant peak corresponding to 3-hydroxypropionyl-CoA (MW 841). This result demonstrates that the 3-hydroxypropionyl-CoA dehydratase activity catalyzes reaction #1.

To detect the conversion of 3-hydroxypropionyl-CoA into acrylyl-CoA, reaction #2 was carried out in 80% $D_2O$. The reaction #2 sample containing the cell extract from cells transfected with the 3-hydroxypropionyl-CoA dehydratase-encoding plasmid revealed incorporation of deuterium in the 3-hydroxypropionyl-CoA molecule. This result indicates that the 3-hydroxypropionyl-CoA dehydratase enzyme catalyzes reaction #2. In addition, the results from both #1 and #2 reactions indicate that the 3-hydroxypropionyl-CoA dehydratase enzyme can catalyze the 3-hydroxypropinyl-CoA ←→ acrylyl-CoA reaction in both directions. It is noted that for both the #1 and #2 reactions, a peak was observed at MW 811, due to leftover acetyl-CoA from the synthesis of 3-hydroxypropionyl-CoA from 3-hydroxypropionate and acetyl-CoA.

The assays assessing conversion of crotonyl-CoA into 3-hydroxybutyryl-CoA also were carried out in 80% $D_2O$. In reaction #3, the control sample exhibited a dominant peak at a molecular weight corresponding to crotonyl-CoA (MW 837). This result indicated that the crotonyl-CoA was not converted into other products. The reaction #3 sample containing the cell extract from cells transfected with the 3-hydroxypropionyl-CoA dehydratase-encoding plasmid exhibited a diffuse group of peaks corresponding to

deuterated 3-hydroxybutyryl-CoA (MW 855 to MW 857). This result demonstrates that the 3-hydroxypropionyl-CoA dehydratase activity catalyzes reaction #3.

A series of control reactions were performed to confirm the specificity of the 3-hydroxypropionyl-CoA dehydratase. Lactyl-CoA (1 mM) was added to the reaction mixture containing 100 mM Tris (pH 7.0) both in the presence and the absence of the 3-hydroxypropionyl-CoA dehydratase. In both cases, the dominant peak observed had a molecular weight corresponding to lactyl-CoA (MW 841). This result indicates that lactyl-CoA is not affected by the presence of 3-hydroxypropionyl-CoA dehydratase activity even in the presence of $D_2O$ meaning that the 3-hydroxypropionyl-CoA dehydratase enzyme does not attach a hydroxyl group at the alpha carbon position. The presence of 3-hydroxypropionyl-CoA in an 80% $D_2O$ reaction mixture resulted in a shift upon addition of the 3-hydroxypropionyl-CoA dehydratase activity. In the absence of 3-hydroxypropionyl-CoA dehydratase activity, a peak corresponding to 3-hydroxypropionyl-CoA was observed in addition to a peak of MW 811. The MW 811 peak was due to leftover acetyl-CoA from the synthesis of 3-hydroxypropionyl-CoA. In the presence of 3-hydroxypropionyl-CoA dehydratase activity, a peak corresponding to deuterated 3-hydroxypropionyl-CoA was observed (MW 842) due to exchange of a hydroxyl group during the conversion of 3-hydroxypropionyl-CoA to acrylyl-CoA and visa-versa. These control reactions demonstrate that the 3-hydroxypropionyl-CoA dehydratase enzyme is active on 3-hydroxypropionyl-CoA and not active on lactyl-CoA. In addition, these results demonstrate that the product of the acrylyl-CoA reaction is 3-hydroxypropionyl-CoA not lactyl-CoA.

## Example 4 - Construction of operon #1

The following operon was constructed and can be used to produce 3-HP in *E. coli* (Figure 34). Briefly, the operon was cloned into a pET-11a expression vector under the control of a T7 promoter (Novagen, Madison, WI). The pET-11a expression vector is a 5677 bp plasmid that uses the ATG sequence of an *NdeI* restriction site as a start codon for inserted downstream sequences.

Nucleic acid molecules encoding a CoA transferase and a lactyl-CoA dehydratase were amplified from *Megasphaera elsdenii* genomic DNA by PCR. Two primers were

used to amplify the CoA transferase-encoding sequence (OSNBpctF 5'-GGGAATTCC-ATATGAGAAAAGTAGAAATCATTACAGCTG-3', SEQ ID NO:108 and OSCTE-2 5'-GAGAGTATACACAGTTTTCACCTCCTTTACAGCAGAGAT-3', SEQ ID NO:109), and two primers were used to amplify the lactyl-CoA dehydratase-encoding sequence (OSCTE-1 5'-ATCTCTGCTGTAAAGGAGGTGAAAACTGTGTATACT-CTC-3', SEQ ID NO:110 and OSEBH-2 5'-ACGTTGATCTCCTTGTACATT-AGAGGATTTCCGAGAAAGC-3', SEQ ID NO:111). A nucleic acid molecule encoding a 3-hydroxypropionyl-CoA dehydratase was amplified from *Chloroflexus aurantiacus* genomic DNA of by PCR using two primers (OSEBH-1 5'-GCTTTCTCGG-AAATCCTCTAATGTACAAGGAGATCAACGT-3', SEQ ID NO:112 and OSHBR 5'-CGACGGATCCTCAACGACCACTGAAGTTGG-3', SEQ ID NO:113).

PCR was conducted in a Perkin Elmer 2400 Thermocycler using 100 ng of genomic DNA and a mix of rTth polymerase (Applied Biosystems; Foster City, CA) and Pfu Turbo polymerase (Stratagene; La Jolla, CA) in 8:1 ratio. The polymerase mix ensured higher fidelity of the PCR reaction. The following PCR conditions were used: initial denaturation step of 94°C for 2 minutes; 20 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 68°C for 2 minutes; and a final extension at 68°C for 5 minutes. The obtained PCR products were gel purified using a Qiagen Gel Extraction Kit (Qiagen, Inc.; Valencia, CA).

The CoA transferase, lactyl-CoA dehydratase (E1, E2 α subunit, and E2 β subunit), and 3-hydroxypropionyl-CoA dehydratase PCR products were assembled using PCR. The OSCTE-1 and OSCTE-2 primers as well as the OSEBH-1 and OSEBH-2 primers were complementary to each other. Thus, the complementary DNA ends could anneal to each other during the PCR reaction extending the DNA in both direction. To ensure the efficiency of the assembly, two end primers (OSNBpctF and OSHBR) were added to the assembly PCR mixture, which contained 100 ng of each PCR product (i.e., the PCR products from the CoA-transferase, lactyl-CoA dehydratase, and 3-hydroxypropionyl-CoA dehydratase reactions) as well as the rTth polymerase/Pfu Turbo polymerase mix described above. The following PCR conditions were used to assemble the products: 94°C for 1 minute; 25 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 68°C for 6 minutes; and a final extension at 68°C for 7 minutes. The assembled PCR

product was gel purified and digested with restriction enzymes (*NdeI* and *BamHI*). The sites for these restriction enzymes were introduced into the assembled PCR product using the OSNBpctF (*NdeI*) and OSHBR (*BamHI*) primers. The digested PCR product was heated at 80°C for 30 minutes to inactive the restriction enzymes and used directly for ligation into pET-11a vector.

The pET-11a vector was digested with *NdeI* and *BamHI* restriction enzymes, gel purified using a Qiagen Gel Extraction kit, treated with shrimp alkaline phosphatase (Roche Molecular Biochemicals; Indianapolis, IN) and used in a ligation reaction with the assembled PCR product. The ligation was performed at 16°C overnight using T4 ligase (Roche Molecular Biochemicals; Indianapolis, IN). The resulting ligation reaction was transformed into NovaBlue chemically competent cells (Novagen; Madison, WI) using a heat-shock method. Once heat shocked, the cells were plated on LB plates supplemented with 50 μg/mL carbenicillin. The plasmid DNA was purified from individual colonies using a QiaPrep Spin Miniprep Kit (Qiagen Inc., Valencia, CA) and analyzed by digestion with *NdeI* and *BamHI* restriction enzymes.

**Example 5 - Construction of operon #2**

The following operon was constructed and can be used to produce 3-HP in *E. coli* (Figure 35A and B). Nucleic acid molecules encoding a CoA transferase and a lactyl-CoA dehydratase were amplified from *Megasphaera elsdenii* genomic DNA by PCR. Two primers were used to amplify the CoA transferase-encoding sequence (OSNBpctF and OSCTE-2), and two primers were used to amplify the lactyl-CoA dehydratase-encoding sequence (OSCTE-1 and OSNBe1R 5'-CGACGGATCCTTAGAGGATTT-CCGAGAAAGC-3', SEQ ID NO:114). A nucleic acid molecule encoding a 3-hydroxypropionyl-CoA dehydratase was amplified from *Chloroflexus aurantiacus* genomic DNA of by PCR using two primers (OSXNhF 5'-GGTGTCT-AGAGACAGTCCTGTCGTTTATGTAGAAGGAG-3', SEQ ID NO:115 and OSXNhR 5'-GGGAATTCCATATGCGTAACTTCCTCCTGCTATCAACGACCACTGAA-GTTGG-3', SEQ ID NO:116).

PCR was conducted in a Perkin Elmer 2400 Thermocycler using 100 ng of genomic DNA and a mix of rTth polymerase (Applied Biosystems; Foster City, CA) and

Pfu Turbo polymerase (Stratagene; La Jolla, CA) in 8:1 ratio. The polymerase mix ensured higher fidelity of the PCR reaction. The following PCR conditions were used: initial denaturation step of 94°C for 2 minutes; 20 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 68°C for 2 minutes; and a final extension at 68°C for 5 minutes. The obtained PCR products were gel purified using a Qiagen Gel Extraction Kit (Qiagen, Inc.; Valencia, CA).

The CoA transferase and lactyl-CoA dehydratase (E1, E2 α subunit, and E2 β subunit) PCR products were assembled using PCR. The OSCTE-1 and OSCTE-2 primers were complementary to each other. Thus, the 22 nucleotides at the end of the CoA transferase sequence and the 22 nucleotides at the beginning of the lactyl-CoA dehydratase could anneal to each other during the PCR reaction extending the DNA in both direction. To ensure the efficiency of the assembly, two end primers (OSNBpctF and OSNBelR) were added to the assembly PCR mixture, which contained 100 ng of the CoA transferase PCR product, 100 ng of lactyl-CoA dehydratase PCR product, and the rTth polymerase/Pfu Turbo polymerase mix described above. The following PCR conditions were used to assemble the products: 94°C for 1 minute; 20 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 68°C for 5 minutes; and a final extension at 68°C for 6 minutes.

The assembled PCR product was gel purified and digested with restriction enzymes (*NdeI* and *BamHI*). The sites for these restriction enzymes were introduced into the assembled PCR product using the OSNBpctF (*NdeI*) and OSNBelR (*BamHI*) primers. The digested PCR product was heated at 80°C for 30 minutes to inactive the restriction enzymes and used directly for ligation into a pET-11a vector.

The pET-11a vector was digested with *NdeI* and *BamHI* restriction enzymes, gel purified using a Qiagen Gel Extraction kit, treated with shrimp alkaline phosphatase (Roche Molecular Biochemicals; Indianapolis, IN) and used in a ligation reaction with the assembled PCR product. The ligation was performed at 16°C overnight using T4 ligase (Roche Molecular Biochemicals; Indianapolis, IN). The resulting ligation reaction was transformed into NovaBlue chemically competent cells (Novagen; Madison, WI) using a heat-shock method. Once heat shocked, the cells were plated on LB plates supplemented with 50 μg/mL carbenicillin. The plasmid DNA was purified from individual colonies

using a QiaPrep Spin Miniprep Kit (Qiagen Inc., Valencia, CA) and analyzed by digestion with *NdeI* and *BamHI* restriction enzymes. The digest revealed that the DNA fragment containing CoA transferase-encoding and lactyl-CoA dehydratase-encoding sequences was cloned into the pET-11a vector.

The plasmid carrying the CoA transferase-encoding and lactyl-CoA dehydratase-encoding sequences (pTD) was digested with *XbaI* and *NdeI* restriction enzymes, gel purified, and used for cloning the 3-hydroxypropionyl-CoA dehydratase-encoding product upstream of the CoA transferase-encoding sequence. Since this *XbaI* and *NdeI* digest eliminated a ribosome-binding site (RBS) from the pET-11a vector, a new homologous RBS was cloned into the plasmid together with the 3-hydroxypropionyl-CoA dehydratase-encoding product. Briefly, the 3-hydroxypropionyl-CoA dehydratase-encoding PCR product was digested with *XbaI* and *NdeI* restriction enzymes, heated at 65°C for 30 minutes to inactivate the restriction enzymes, and ligated into pTD. The ligation mixture was transformed into chemically competent NovaBlue cells (Novagen) that then were plated on LB plates supplemented with 50 μg/mL carbenicillin.

Individual colonies were selected, and the plasmid DNA obtained using a Qiagen Spin Miniprep Kit. The obtained plasmids were digested with *XbaI* and *NdeI* restriction enzymes and analyzed by gel electrophoresis. pTD plasmids containing the inserted 3-hydroxypropionyl-CoA dehydratase-encoding PCR product were named pHTD. While expression of the lactyl-CoA hydratase, CoA transferase, and 3-hydroxypropionyl-CoA dehydratase sequences from pHTD was directed by a single T7 promoter, each coding sequence had an individual RBS upstream of their start codon.

To ensure the correct assembly and cloning of the lactyl-CoA hydratase, CoA transferase, and 3-hydroxypropionyl-CoA dehydratase sequences into one operon, both ends of the operon and all junctions between the coding sequences were sequenced. This DNA analysis revealed that the operon was assembled correctly.

The pHTD plasmid was transformed into BL21(DE3) cells to study the expression of the encoded sequences.

**Example 6 - Construction of operons #3 and #4**

Operon #3 (Figure 36A and B) and operon #4 (Figure 37A and B) each position the E1 activator at the end of the operon. Operon #3 contains a RBS between the 3-hydroxypropionyl-CoA dehydratase-encoding sequence and the E1 activator-encoding sequence. In operon #4, however, the stop codon of the 3-hydroxypropionyl-CoA dehydratase-encoding sequence is fused with the start codon of the E1 activator-encoding sequence as follows: TAGTG. The absence of the RBS in operon #4 can decrease the level of E1 activator expression.

To construct operon #3, nucleic acid molecules encoding a CoA transferase and a lactyl-CoA dehydratase were amplified from *Megasphaera elsdenii* genomic DNA by PCR. Two primers were used to amplify the CoA transferase-encoding sequence (OSNBpctF and OSHTR 5'-ACGTTGATCTCCTTCTACATTATTTTTTCAGT-CCCATG-3', SEQ ID NO:117), two primers were used to amplify the E2 α and β subunits of the lactyl-CoA dehydratase-encoding sequence (OSEIIXNF 5'-GGTGTCTAGAGTCAAAGGAGAGAACAAAATCATGAGTG-3', SEQ ID NO:118 and OSEIIXNR 5'-GGGAATTCCATATGCGTAACTTCCTCCTGCTATTAGAGGA-TTTCCGAGAAAGC-3', SEQ ID NO:119), and two primers were used to amplify the E1 activator of the lactyl-CoA dehydratase-encoding sequence (OSHrEIF 5'-TCAGTG-GTCGTTGATCACGCTATAAAGAAAGGTGAAAACTGTGTATACTCTC-3', SEQ ID NO:120 and OSEIBR 5'-CGACGGATCCCTTCCTTGGAGCTCATGCTTTC-3', SEQ ID NO:121). A nucleic acid molecule encoding a 3-hydroxypropionyl-CoA dehydratase was amplified from *Chloroflexus aurantiacus* genomic DNA of by PCR using two primers (OSTHF 5'-CATGGGACTGAAAAAATAATGTAGAAGGAGAT-CAACGT-3', SEQ ID NO:122 and OSEIrHR 5'-GAGAGTATACACAGTTTTCA-CCTTTCTTTATAGCGTGATCAACGACCACTGA-3', SEQ ID NO:123).

PCR was conducted in a Perkin Elmer 2400 Thermocycler using 100 ng of genomic DNA and a mix of rTth polymerase (Applied Biosystems; Foster City, CA) and Pfu Turbo polymerase (Stratagene; La Jolla, CA) in 8:1 ratio. The polymerase mix ensured higher fidelity of the PCR reaction. The following PCR conditions were used: initial denaturation step of 94°C for 2 minutes; 20 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 68°C for 2 minutes; and a final extension at 68°C for 5 minutes. The

obtained PCR products were gel purified using a Qiagen Gel Extraction Kit (Qiagen, Inc.; Valencia, CA).

The 3-hydroxypropionyl-CoA dehydratase and E1 activator PCR products were assembled using PCR. The OSHrE1F and OSEIrHR primers were complementary to each other. Thus, the primers could anneal to each other during the PCR reaction extending the DNA in both direction. To ensure the efficiency of the assembly, two end primers (OSTHF and OSE1BR) were added to the assembly PCR mixture, which contained 100 ng of the 3-hydroxypropionyl-CoA dehydratase PCR product, 100 ng of E1 activator PCR product, and the rTth polymerase/Pfu Turbo polymerase mix described above. The following PCR conditions were used to assemble the products: 94°C for 1 minute; 20 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 68°C for 1.5 minutes; and a final extension at 68°C for 5 minutes.

The assembled PCR product was gel purified and used in a second assembly PCR with gel purified the CoA transferase PCR product. The OSTHF and OSHTR primers were complementary to each other. Thus, the complementary DNA ends could anneal to each other during the PCR reaction extending the DNA in both direction. To ensure the efficiency of the assembly, two end primers (OSNBpctF and OSEIBR) were added to the second assembly PCR mixture, which contained 100 ng of the purified 3-hydroxypropionyl-CoA dehydratase/EI PCR assembly, 100 ng of the purified CoA transferase PCR product, and the polymerase mix described above. The following PCR conditions were used to assemble the products: 94°C for 1 minute; 20 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 68°C for 3 minutes; and a final extension at 68°C for 5 minutes.

The assembled PCR product was gel purified and digested with *NdeI* and *BamHI* restriction enzymes. The sites for these restriction enzymes were introduced into the assembled PCR products with the OSNBpctF (*NdeI*) and OSEIBR (*BamHI*) primers. The digested PCR product was heated at 80°C for 30 minutes to inactive the restriction enzymes and used directly for ligation into a pET11a vector.

The pET-11a vector was digested with *NdeI* and *BamHI* restriction enzymes, gel purified using a Qiagen Gel Extraction kit, treated with shrimp alkaline phosphatase (Roche Molecular Biochemicals; Indianapolis, IN) and used in a ligation reaction with the

assembled PCR product. The ligation was performed at 16°C overnight using T4 ligase (Roche Molecular Biochemicals; Indianapolis, IN). The resulting ligation reaction was transformed into NovaBlue chemically competent cells (Novagen; Madison, WI) using a heat-shock method. Once heat shocked, the cells were plated on LB plates supplemented with 50 μg/mL carbenicillin. The plasmid DNA was purified from individual colonies using a QiaPrep Spin Miniprep Kit (Qiagen Inc.; Valencia, CA). The resulting plasmids carrying the CoA transferase, 3-hydroxypropionyl-CoA dehydratase, and EI activator sequences (pTHrEI) were digested with *XbaI* and *NdeI*, purified using gel electrophoresis and a Qiagen Gel Extraction kit, and used as a vector for cloning of the E2 α subunit/E2 β subunit PCR product.

The E2 α subunit/E2 β subunit PCR product was digested with the same enzymes and ligated into the pTHrEI vector. The ligation reaction was performed at 16°C overnight using T4 ligase (Roche Molecular Biochemicals; Indianapolis, IN). The ligation mixture was transformed into chemically competent NovaBlue cells (Novagen) that then were plated on LB plates supplemented with 50 μg/mL carbenicillin. The plasmid DNA was purified from individual colonies using a QiaPrep Spin Miniprep Kit (Qiagen Inc., Valencia, CA) and digested with *XbaI* and *NdeI* restriction enzymes for gel electrophoresis analysis. The resulting plasmids carrying the constructed operon #3 (pEIITHrEI) were transformed into BL21(DE3) cells to study the expression of the cloned sequences. Electrospray mass spectrometry assay confirmed that extracts from these cells have CoA transferase activity and 3-hydroxypropionyl-CoA dehydratase activity. Similar assays are used to confirm that extracts from these cells also have lactyl-CoA dehydratase activity.

To construct operon #4, nucleic acid molecules encoding a CoA transferase and a lactyl-CoA dehydratase were amplified from *Megasphaera elsdenii* genomic DNA by PCR. Two primers were used to amplify the CoA transferase-encoding sequence (OSNBpctF and OSHTR), two primers were used to amplify the E2 α and β subunits of the lactyl-CoA dehydratase-encoding sequence (OSEIIXNF and OSEIIXNR), and two primers were used to amplify the E1 activator of the lactyl-CoA dehydratase-encoding sequence (OSHEIF 5'-CCAACTTCAGTGGTCGTTAGTGAAAACTGTGTAT-ACTCTC-3', SEQ ID NO:124 and OSEIBR). A nucleic acid molecule encoding a 3-

hydroxypropionyl-CoA dehydratase was amplified from *Chloroflexus aurantiacus* genomic DNA of by PCR using two primers (OSTHF and OSEIHR 5'-GAGAGTATACACAGTTTTCACTAACGACCACTGAAGTTGG-3', SEQ ID NO:125).

PCR was conducted in a Perkin Elmer 2400 Thermocycler using 100 ng of genomic DNA and a mix of rTth polymerase (Applied Biosystems; Foster City, CA) and Pfu Turbo polymerase (Stratagene; La Jolla, CA) in 8:1 ratio. The polymerase mix ensured higher fidelity of the PCR reaction. The following PCR conditions were used: initial denaturation step of 94°C for 2 minutes; 20 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 68°C for 2 minutes; and a final extension at 68°C for 5 minutes. The obtained PCR products were gel purified using a Qiagen Gel Extraction Kit (Qiagen, Inc.; Valencia, CA).

The 3-hydroxypropionyl-CoA dehydratase and E1 activator PCR products were assembled using PCR. The OSHEIF and OSEIHR primers were complementary to each other. Thus, the primers could anneal to each other during the PCR reaction extending the DNA in both direction. To ensure the efficiency of the assembly, two end primers (OSTHF and OSE1BR) were added to the assembly PCR mixture, which contained 100 ng of the 3-hydroxypropionyl-CoA dehydratase PCR product, 100 ng of E1 activator PCR product, and the rTth polymerase/Pfu Turbo polymerase mix described above. The following PCR conditions were used to assemble the products: 94°C for 1 minute; 20 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 68°C for 1.5 minutes; and a final extension at 68°C for 5 minutes.

The assembled PCR product was gel purified and used in a second assembly PCR with gel purified the CoA transferase PCR product. The OSTHF and OSHTR primers were complementary to each other. Thus, the complementary DNA ends could anneal to each other during the PCR reaction extending the DNA in both direction. To ensure the efficiency of the assembly, two end primers (OSNBpctF and OSEIBR) were added to the second assembly PCR mixture, which contained 100 ng of the purified 3-hydroxypropionyl-CoA dehydratase/EI PCR assembly, 100 ng of the purified CoA transferase PCR product, and the polymerase mix described above. The following PCR conditions were used to assemble the products: 94°C for 1 minute; 20 cycles of 94°C for

30 seconds, 54°C for 30 seconds, and 68°C for 3 minutes; and a final extension at 68°C for 5 minutes.

The assembled PCR product was gel purified and digested with *NdeI* and *BamHI* restriction enzymes. The sites for these restriction enzymes were introduced into the assembled PCR products with the OSNBpctF (*NdeI*) and OSEIBR (*BamHI*) primers. The digested PCR product was heated at 80°C for 30 minutes to inactive the restriction enzymes and used directly for ligation into a pET11a vector.

The pET-11a vector was digested with *NdeI* and *BamHI* restriction enzymes, gel purified using a Qiagen Gel Extraction kit, treated with shrimp alkaline phosphatase (Roche Molecular Biochemicals; Indianapolis, IN) and used in a ligation reaction with the assembled PCR product. The ligation was performed at 16°C overnight using T4 ligase (Roche Molecular Biochemicals; Indianapolis, IN). The resulting ligation reaction was transformed into NovaBlue chemically competent cells (Novagen; Madison, WI) using a heat-shock method. Once shocked, the cells were plated on LB plates supplemented with 50 μg/mL carbenicillin. The plasmid DNA was purified from individual colonies using a QiaPrep Spin Miniprep Kit (Qiagen Inc., Valencia, CA). The resulting plasmids carrying the CoA transferase, 3-hydroxypropionyl-CoA dehydratase, and EI activator sequences (pTHE1) were digested with *XbaI* and *NdeI*, purified using gel electrophoresis and a Qiagen Gel Extraction kit, and used as a vector for cloning of the E2 α subunit/E2 β subunit PCR product.

The E2 α subunit/E2 β subunit PCR product was digested with the same enzymes and ligated into the pTHE1 vector. The ligation reaction was performed at 16°C overnight using T4 ligase (Roche Molecular Biochemicals, Indianapolis, IN). The ligation mixture was transformed into chemically competent NovaBlue cells (Novagen) that then were plated on LB plates supplemented with 50 μg/mL carbenicillin. The plasmid DNA was purified from individual colonies using a QiaPrep Spin Miniprep Kit (Qiagen Inc., Valencia, CA) and digested with *XbaI* and *NdeI* restriction enzymes for gel electrophoresis analysis. The resulting plasmids carrying the constructed operon #4 (pEIITHEI) were transformed into BL21(DE3) cells to study the expression of the cloned sequences. Electrospray mass spectrometry assays confirmed that extracts from these cells have CoA transferase activity and 3-hydroxypropionyl-CoA dehydratase activity.

Similar assays are used to confirm that extracts from these cells also have lactyl-CoA dehydratase activity.

*E. coli* plasmid pEIITHrEI carrying a synthetic 3-HP operon was digested with *Nru*I, *Xba*I and *Bam*HI restriction enzymes, *Xba*I-*Bam*HI DNA fragment was gel purified with Quagen Gel Extraction Kit (Qiagen, Inc., Valencia CA) and used for further cloning into Bacillu vector pWH1520 (MoBiTec BmBH, Gottingen, Germany). Vector pWH1520 was digested with *Spe*I and *Bam*HI restriction enzymes and gel purified with Qiagen Gel Extraction Kit. *The Xba*I-*BamHI* fragment carrying 3-HP operon was ligated into WH1520 vector at 16°C overnight using T4 ligase. The ligation mixture was transformed into chemically competent TOP 10 cells and plated on LB plates supplemented with 50 μg/ml carbenicillin. One clone named *B. megaterium* (pBPO26) was used for assays of CoA-transferase and CoA-hydratase activities. The assays were performed as described above for E. Coli. The enzymatic activity was 5 U/mg and 13 U/mg respectively.

**Example 7 - Construction of a two plasmid system**

The following constructs were constructed and can be used to produce 3-HP in *E. coli* (Figure 38A and B). Nucleic acid molecules encoding a CoA transferase and a lactyl-CoA dehydratase were amplified from *Megasphaera elsdenii* genomic DNA by PCR. Two primers were used to amplify the CoA transferase-encoding sequence (OSNBpctF and OSHTR), two primers were used to amplify the E2 α and β subunits of the lactyl-CoA dehydratase-encoding sequence (OSEIIXNF and OSEIIXNR), and two primers were used to amplify the E1 activator of the lactyl-CoA dehydratase-encoding sequence (E1PROF 5'-GTCGCAGAATTCCCATCAATCGCAGCAATCCCAAC-3', SEQ ID NO:126 and E1PROR 5'-TAACATGGTACCGACAGAAGCGGACCAGCA-AACGA-3', SEQ ID NO:127). A nucleic acid molecule encoding a 3-hydroxypropionyl-CoA dehydratase was amplified from *Chloroflexus aurantiacus* genomic DNA of by PCR using two primers (OSTHF and OSHBR 5'-CGACGGATCCTCAACGACCA-CTGAAGTTGG-3', SEQ ID NO:128).

PCR was conducted in a Perkin Elmer 2400 Thermocycler using 100 ng of genomic DNA and a mix of rTth polymerase (Applied Biosystems; Foster City, CA) and

Pfu Turbo polymerase (Stratagene; La Jolla, CA) in 8:1 ratio. The polymerase mix ensured higher fidelity of the PCR reaction. The following PCR conditions were used: initial denaturation step of 94°C for 2 minutes; 20 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 68°C for 2 minutes; and a final extension at 68°C for 5 minutes. The obtained PCR products were gel purified using a Qiagen Gel Extraction Kit (Qiagen, Inc.; Valencia, CA).

The CoA transferase PCR product and the 3-hydroxypropionyl-CoA dehydratase PCR product were assembled using PCR. The OSTHF and OSHTR primers were complementary to each other. Thus, the complementary DNA ends could anneal to each other during the PCR reaction extending the DNA in both direction. To ensure the efficiency of the assembly, two end primers (OSNBpctF and OSHBR) were added to the assembly PCR mixture, which contained 100 ng of the purified CoA transferase PCR product, 100 ng of the purified 3-hydroxypropionyl-CoA dehydratase PCR product, and the polymerase mix described above. The following PCR conditions were used to assemble the products: 94°C for 1 minute; 20 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 68°C for 2.5 minutes; and a final extension at 68°C for 5 minutes.

The assembled PCR product was gel purified and digested with *NdeI* and *BamHI* restriction enzymes. The sites for these restriction enzymes were introduced into the assembled PCR products with the OSNBpctF (*NdeI*) and OSHBR (*BamHI*) primers. The digested PCR product was heated at 80°C for 30 minutes to inactive the restriction enzymes and used directly for ligation into a pET11a vector.

The pET-11a vector was digested with *NdeI* and *BamHI* restriction enzymes, gel purified using a Qiagen Gel Extraction kit, treated with shrimp alkaline phosphatase (Roche Molecular Biochemicals; Indianapolis, IN) and used in a ligation reaction with the assembled PCR product. The ligation was performed at 16°C overnight using T4 ligase (Roche Molecular Biochemicals; Indianapolis, IN). The resulting ligation reaction was transformed into NovaBlue chemically competent cells (Novagen; Madison, WI) using a heat-shock method. Once shocked, the cells were plated on LB plates supplemented with 50 μg/mL carbenicillin. The plasmid DNA was purified from individual colonies using a QiaPrep Spin Miniprep Kit (Qiagen Inc.; Valencia, CA) and digested with *NdeI* and *BamHI* restriction enzymes for gel electrophoresis analysis. The resulting plasmids

carrying the CoA transferase and 3-hydroxypropionyl-CoA dehydratase (pTH) were digested with *XbaI* and *NdeI*, purified using gel electrophoresis and a Qiagen Gel Extraction kit, and used as a vector for cloning of the E2 α subunit/E2 β subunit PCR product.

The E2 α subunit/E2 β subunit PCR product digested with the same enzymes was ligated into the pTH vector. The ligation reaction was performed at 16°C overnight using T4 ligase (Roche Molecular Biochemicals; Indianapolis, IN). The ligation mixture was transformed into chemically competent NovaBlue cells (Novagen) that then were plated on LB plates supplemented with 50 μg/mL carbenicillin. The plasmid DNA was purified from individual colonies using a QiaPrep Spin Miniprep Kit (Qiagen Inc.; Valencia, CA) and digested with *XbaI* and *NdeI* restriction enzymes for gel electrophoresis analysis. The resulting plasmids carrying the E2 α and β subunits of the lactyl-CoA dehydratase, the CoA transferase, and the 3-hydroxypropionyl-CoA dehydratase (pEIITH) were transformed into BL21(DE3) cells to study the expression of the cloned sequences.

The gel purified E1 activator PCR product was digested with *EcoRI* and *KpnI* restriction enzymes, heated at 65°C for 30 minutes, and ligated into a vector (pPROLar.A) that was digested with *EcoRI* and *KpnI* restriction enzymes, gel purified using Qiagen Gel Extraction kit, and treated with shrimp alkaline phosphatase (Roche Molecular Biochemicals; Indianapolis, IN). The ligation was performed at 16°C overnight using T4 ligase (Roche Molecular Biochemicals; Indianapolis, IN). The resulting ligation reaction was transformed into DH10B electro-competent cells (Gibco Life Technologies; Gaithersburg, MD) using electroporation. Once electroporated, the cells were plated on LB plates supplemented with 25 μg/mL kanamycin. The plasmid DNA was purified from individual colonies using a QiaPrep Spin Miniprep Kit (Qiagen Inc., Valencia, CA) and digested with *EcoRI* and *KpnI* restriction enzymes for gel electrophoresis analysis. The resulting plasmids carrying the E1 activator (pPROEI) are transformed into BL21(DE3) cells to study the expression of the cloned sequence.

The pPROEI and pEIITH plasmids are compatible plasmids that can be used in the same bacterial host cell. In addition, the expression from the pPROEI and pEIITH plasmids can be induced at different levels using IPTG and arabinose, allowing for the fine-tuning of the expression of the cloned sequences.

**Example 8 – Production of 3-HP**

3-HP was produced using recombinant *E. coli* in a small-scale batch fermentation reaction. The construction of strain ALS848 (also designated as TA3476 (*J. Bacteriol.*, 143:1081-1085(1980))) that carried inducible T7 RNA polymerase was performed using λDE3 lysogenization kit (Novagen, Madison, WI) according to the manufacture's instructions. The constructed strain was designated ALS484(DE3). Strain ALS484(DE3) was transformed with pEIITHrEI plasmid using standard electroporation techniques. The transformants were selected on LB/carbenicillin (50 μg/mL) plates. A single colony was used to inoculate 4 mL culture in a 15 mL culture tube. Strain ALS484(DE3) strain carrying vector pET11a was used as a control. The cells were grown at 37°C and 250 rpm in an Innova 4230 Incubator Shaker (New Brunswick Scientific, Edison, NJ) for eight to nine hours. This culture (3 mL) was used to start an anaerobic fermentation. Two 100 mL anaerobic cultures of ALS(DE3)/pET11a and ALS(DE3)pEIITHrEI were grown in serum bottles using LB media supplemented with 0.4% glucose, 50 μg/mL carbenicillin, and 100 mM MOPS. The cultures were grown overnight at 37°C without shaking. The overnight grown cultures were sub-cultured in serum bottles using fresh LB media supplemented with 0.4% glucose, 50 μg/mL carbenicillin, and 100 mM MOPS. The starting OD(600) of these cultures was adjusted to 0.3. These serum bottles were incubated at 37°C without shaking. After one hour of incubation, the cultures were induced with 100 μM IPTG. A 3 mL sample was taken from each of the serum bottles at 30 minutes, 1 hour, 2 hours, 4 hours, 6 hours, 8 hours, and 24 hours. The samples were transferred into two properly labeled 2 mL microcentrifuge tubes, each containing 1.5 mL sample. The samples were spun down in a microcentrifuge centrifuge at 14000 g for 3 minutes. The supernatant was passed through a 0.45 μ syringe filter, and the resulting filtrate was stored at -20°C until further analysis. The formation of fermentation products, mainly lactate and 3-hydroxypropionate, was measured by detecting derivatized CoA esters of lactate and 3-HP using LC/MS.

The following methods were performed to convert lactate and 3-HP into their respective CoA esters. Briefly, the filtrates were mixed with CoA-reaction buffer (200 mM potassium phosphate buffer, 2 mM acetyl-CoA, and 0.1 mg/mL purified transferase) in 1:1 ratio. The reaction was allowed to proceed for 20 minutes at room temperature.

The reaction was stopped by adding 1 volume of 10% TFA. The sample was purified using Sep Pak Vac columns (Waters). The column was conditioned with methanol and washed two times with 0.1% TFA. The sample was then applied to the column, and the column was washed two more times with 0.1% TFA. The sample was eluted with 40% acetonitrile, 0.1% TFA. The acetonitrile was removed from the sample by vacuum centrifugation. The samples were then analyzed by LC/MS.

Analysis of the standard CoA/CoA thioester mixtures and the CoA thioester mixtures derived from fermentation broths were carried out using a Waters/Micromass ZQ LC/MS instrument which had a Waters 2690 liquid chromatograph with a Waters 996 Photo-Diode Array (PDA) absorbance monitor placed in series between the chromatograph and the single quadrupole mass spectrometer. LC separations were made using a 4.6 x 150 mm YMC ODS-AQ (3 μm particles, 120 Å pores) reversed-phase chromatography column at room temperature. Two gradient elution systems were developed using different mobile phases for the separation of the CoA esters. These two systems are summarized in Table 3. Elution system 1 was developed to provide the most rapid and efficient separation of the five-component CoA/CoA thioester mixture (CoA, acetyl-CoA, lactyl-CoA, acrylyl-CoA, and propionyl-CoA), while elution system 2 was developed to provide baseline separation of the structurally isomeric esters lactyl-CoA and 3HP-CoA in addition to separation of the remaining esters listed above. In all cases, the flow rate was 0.250 mL/minute, and photodiode array UV absorbance was monitored from 200 nm to 400 nm. All parameters of the electrospray MS system were optimized and selected based on generation of protonated molecular ions ($[M + H]^+$) of the analytes of interest and production of characteristic fragment ions. The following instrumental parameters were used for ESI-MS detection of CoA and organic acid-CoA thioesters in the positive ion mode: Capillary: 4.0 V; Cone: 56 V; Extractor: 1 V; RF lens: 0 V; Source temperature: 100°C; Desolvation temperature: 300°C; Desolvation gas: 500 L/hour; Cone gas: 40 L/hour; Low mass resolution: 13.0; High mass resolution: 14.5; Ion energy: 0.5; Multiplier: 650. Uncertainties for reported mass/charge ratios (*m/z*) and molecular masses are ± 0.01%. Table 3 provides a summary of gradient elution systems for the separation of organic acid-Coenzyme A thioesters.

Table 3

| System | Buffer A | Buffer B | Gradient | |
|---|---|---|---|---|
| | | | Time | Percent B |
| 1 | 25 mM ammonium acetate<br>0.5 % acetic acid | ACN<br>0.5 % acetic acid | 0<br>40<br>42<br>47<br>50 | 10<br>40<br>100<br>100<br>10 |
| 2 | 25 mM ammonium acetate<br>10 mM TEA<br>0.5 % acetic acid | ACN<br>0.5 % acetic acid | 0<br>10<br>45<br>50<br>53<br>54 | 10<br>10<br>60<br>100<br>100<br>10 |

The following methods were used to separate the derivatized 3-hydroxypropionyl-CoA, which was formed from 3-HP, from 2-hydroxypropionyl-CoA (i.e., lactyl-CoA), which was formed from lactate. Because these structural isomers have identical masses and mass spectral fragmentation behavior, the separation and identification of these analytes in a mixture depends on their chromatographic separation. While elution system 1 provided excellent separation of the CoA thioesters tested (Figure 46), it was unable to resolve 3-HP-CoA and lactyl-CoA. An alternative LC elution system was developed using ammonium acetate and triethylamine (system 2; Table 3).

The ability of system 2 to separate 3-HP-CoA and lactyl-CoA was tested on a mixture of these two compounds. Comparing the results from a mixture of 3-HP-CoA and lactyl-CoA (Figure 47, Panel A) to the results from lactyl-CoA only (Figure 47, Panel B) revealed that system 2 can separate 3-HP-CoA and lactyl-CoA. The mass spectrum recorded under peak 1 (Figure 47, Panel A insert) was used to identify peak 1 as being a hydroxypropionyl-CoA thioester when compared to Figure 46, Panel C. In addition, comparison of Panels A and B of Figure 47 as well as the mass spectra results

corresponding to each peak revealed that peak 1 corresponds to 3-HP-CoA and peak 2 corresponds to lactyl-CoA.

System 2 was used to confirm that *E. coli* transfected with pEIITHrEI produced 3-HP in that 3-HP-CoA was detected. Specifically, an ion chromatogram for $m/z = 840$ in the analysis of a CoA transferase-treated fermentation broth aliquot collected from a culture of *E. coli* containing pEIITHrEI revealed the presence of 3-HP-CoA (Figure 48, Panel A). The CoA transferase-treated fermentation broth aliquot collected from a culture of *E. coli* lacking pEIITHrEI did not exhibit the peak corresponding to 3-HP-CoA (Figure 48, Panel B). Thus, these results indicate that the pEIITHrEI plasmid directs the expression of polypeptides having propionyl-CoA transferase activity, lactyl-CoA dehydratase activity, and acrylyl-CoA hydratase activity. These results also indicate that expression of these polypeptides leads to the formation of 3-HP.

**Example 9 – Cloning nucleic acid molecules that encode a polypeptide having acetyl CoA carboxylase activity**

Polypeptides having acetyl-CoA carboxylase activity catalyze the first committed step of the fatty acid synthesis by carboxylation of acetyl-CoA to malonyl-CoA. Polypeptides having acetyl-CoA carboxylase activity are also responsible for providing malonyl-CoA for the biosynthesis of very-long-chain fatty acids required for proper cell function. Polypeptides having acetyl-CoA carboxylase activity can be biotin dependent enzymes in which the cofactor biotin is post-translationally attached to a specific lysine residue. The reaction catalyzed by such polypeptides consists of two discrete half reactions. In the first half reaction, biotin is carboxylated by biocarbonate in an ATP-dependent reaction to form carboxybiotin. In the second half reaction, the carboxyl group is transferred to acetyl-CoA to form malonyl-CoA.

Prokaryotic and eukaryotic polypeptides having acetyl-CoA carboxylase activity exist. The prokaryotic polypeptide is a multi-subunit enzyme (four subunits), where each of the subunits is encoded by a different nucleic acid sequence. For example, in *E. coli*, the following genes encode for the four subunits of acetyl-CoA carboxylase:

*acc*A: Acetyl-coenzyme a carboxylase carboxyl transferase subunit alpha (GenBank® accession number M96394)

*acc*B: Biotin carboxyl carrier protein (GenBank® accession number U18997)

*acc*C: Biotin carboxylase (GenBank® accession number U18997)

*acc*D:. Acetyl-coenzyme a carboxylase carboxyl transferase subunit beta (GenBank® accession number M68934)

The eukaryotic polypeptide is a high molecular weight multi-functional enzyme encoded by a single gene. For example, in *Saccharomyces cerevisiae,* the acetyl-CoA carboxylase can have the sequence set forth in GenBank® accession number M92156.

The prokaryotic type acetyl-CoA carboxylase from *E. coli* was overexpressed using T7 promoter vector pFN476 as described elsewhere (Davis *et al. J. Biol. Chem.*, 275:28593-28598 (2000)). The eukaryotic type acetyl-CoA carboxylase gene was amplified from *Saccharomyces cerevisiae* genomic DNA. Two primers were designed to amplify the *acc*1 gene from in *S. cerevisiae* (acc1F 5'-atag*GCGGCCGC*AGGAATGCTGT**ATGAGCGAAGAAAGCTTATT** C-3', SEQ ID NO: 138 where the bold is homologous sequence, the italics is a *Not* I site, the underline is a RBS, and the lowercase is extra; and acc1R 5'-atgctcgcat*CTCGAG***TAG-CTAAATTAAATTACATCAATAGTA**-3', SEQ ID NO: 139 where the bold is homologous sequence, the italics is a *Xho* I site, and the lowercase is extra). The following PCR mix is used to amplify *acc*1 gene 10X *pfu* buffer (10 μL), dNTP (10mM; 2 μL), cDNA (2 μL), acc1F (100 μM; 1 μL), acc1R (100 μM; 1 μL), pfu enzyme (2.5 units/μL; 2 μL), and DI water (82 μL). The following protocol was used to amplify the acc1 gene. After performing PCR, the PCR product was separated on a gel, and the band corresponding to acc1 nucleic acid (about 6.7 Kb) was gel isolated using Qiagen gel isolation kit. The PCR fragment is digested with *Not* I and *Xho* I (New England BioLab) restriction enzymes. The digested PCR fragment is then ligated to pET30a which was restricted with *Not* I and *Xho* I and dephosphorylated with SAP enzyme. The *E.coli* strain DH10B was transformed with 1 μL of the ligation mix, and the cells were recovered in 1 mL of SOC media. The transformed cells were selected on LB/kanamycin (50 μg/μL) plates. Eight single colonies are selected, and PCR was used to screen for the correct insert. The plasmid having correct insert was isolated using Qiagen Spin Mini prep kit.

To obtain a polypeptide having acetyl-CoA carboxylase activity, the plasmid pMSD8 or pET30a/acc1 overexpressing *E. coli* or *S. cerevisiae* acetyl-CoA carboxylase was transformed into Tuner pLacI chemically competent cells (Novagen, Madison, WI). The transformed cells were selected on LB/chloramphenicol (25 μg/mL) plus carbencillin (50 μg/mL) or kanamycin (50 μg/mL).

A crude extract of this strain can be prepared in the following manner. An overnight culture of Tuner pLacI with pMSD8 is subcultured into 200 mL (in one liter baffle culture flask) of fresh M9 media supplemented with 0.4% glucose, 1 μg/mL thiamine, 0.1% casamino acids, and 50 μg/mL carbencillin or 50 μg/mL kanamycin and 25 μg/mL chloramphenicol. The culture is grown at 37°C in a shaker with 250 rpm agitation until it reaches an optical density at 600 nm of about 0.6. IPTG is then added to a final concentration of 100 μM. The culture is then incubated for an additional 3 hours with shaking speed of 250 rpm at 37°C. Cells are harvested by centrifugation at 8000 x g and are washed one time with 0.85% NaCl. The cell pellet was resuspended in a minimal volume of 50 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 100 mM KCl, 2 mM DTT, and 5% glycerol. The cells are lysed by passing them two times through a French Pressure cell at 1000 psig pressure. The cell debris was removed by centrifugation for 20 minutes at 30,000 x g.

The enzyme can be assayed using a method from Davis *et al.* (*J. Biol. Chem.*, 275:28593-28598 (2000)).

**Example 10 – Cloning a nucleic acid molecule that encodes a polypeptide having malonyl-CoA reductase activity from *Chloroflexus auarantiacus***

A polypeptide having malonyl-CoA reductase activity was partially purified from *Chloroflexus auarantiacus* and used to obtained amino acid micro-sequencing results. The amino acid sequencing results were used to identify and clone the nucleic acid that encodes a polypeptide having malonyl-CoA reductase activity.

Biomass required for protein purification was grown in B. Braun BIOSTAT B fermenters (B. Braun Biotech International GmbH, Melsungen, Germany). A glass vessel fitted with a water jacket for heating was used to grow the required biomass. The glass

vessel was connected to its own control unit. The liquid working volume was 4 L, and the fermenter was operated at 55°C with 75 rpm of agitation. Carbon dioxide was occasionally bubbled through the headspace of the fermenter to maintain anaerobic conditions. The pH of the cultures was monitored using a standard pH probe and was maintained between 8.0 and 8.3. The inoculum for the fermenters was grown in two 250 mL bottles in an Innova 4230 Incubator, Shaker (New Brunswick Scientific, Edison, NJ) at 55°C with interior lights. The fermenters were illuminated by three 65 W plant light reflector lamps (General Electric, Cleveland, OH). Each lamp was placed approximately 50 cm away from the glass vessel. The media used for the inoculum and the fermenter culture was as follows per liter: 0.07 g EDTA, 1 mL micronutrient solution, 1 mL $FeCl_3$ solution, 0.06 g $CaSO_4 \cdot 2\ H_2O$, 0.1 g $MgSO_4 \cdot 7\ H_2O$, 0.008 g NaCl, 0.075 g KCl, 0.103 g $KNO_3$, 0.68 g $NaNO_3$, 0.111 g $Na_2HPO_4$, 0.2 g $NH_4Cl$, 1 g yeast extract, 2.5 g casamino acid, 0.5 g Glycyl-Glycine, and 900 mL DI water. The micronutrient solution contained the following per liter: 0.5 mL $H_2SO_4$ (conc.), 2.28 g $MnSO_4 \cdot 7\ H_2O$, 0.5 g $ZnSO_4 \cdot 7\ H_2O$, 0.5 g $H_3BO_3$, 0.025 g $CuSO_4 \cdot 2\ H_2O$, 0.025 g $Na_2MoO_4 \cdot 2\ H_2O$, and 0.045 g $CoCl_2 \cdot 6\ H_2O$. The $FeCl_3$ solution contained 0.2905 g $FeCl_3$ per liter. After adjusting the pH of the media to 8.2 to 8.4, 0.75 g/L $Na_2S \cdot 9H_2O$ was added, the pH was readjusted to 8.2 to 8.4, and the media was filter-sterilized through a 0.22 μ filter.

The fermenter was inoculated with 500 mL of grown culture. The fermentation was stopped, and the biomass was harvested after the cell density was about 0.5 to 0.6 units at 600 nm.

The cells were harvested by centrifugation at 5000 x g (Beckman JLA 8.1000 rotor) at 4°C, washed with 1 volume of ice cold 0.85% NaCl, and centrifuged again. The cell pellet was resuspended in 30 mL of ice cold 100 mM Tris-HCl (pH 7.8) buffer that was supplemented with 2 mM DTT, 5 mM $MgCl_2$, 0.4 mM PEFABLOC (Roche Molecular Biochemicals, Indianapolis, IN), 1% streptomycin sulfate, and 2 tablets of Complete EDTA-free protease inhibitor cocktail (Roche Molecular Biochemicals, Indianapolis, IN). The cell suspension was lysed by passing the suspension, three times, through a 50 mL French Pressure Cell operated at 1600 psi (gauge reading). Cell debris was removed by centrifugation at 30,000 x g (Beckman JA 25.50 rotor). The crude extract was filtered prior to chromotography using a 0.2 μm HT Tuffryn membrane

syringe filter (Pall Corp., Ann Arbor, MI). The protein concentration of the crude extract was 29 mg/mL, which was determined using the BioRad Protein Assay according to the manufacturer's microassay protocol. Bovine gamma globulin was used for the standard curve determination. This assay was based on the Bradford dye-binding procedure (Bradford, *Anal. Biochem.*, 72:248 (1976)).

Before starting the protein purification, the following assay was used to determine the activity of malonyl-CoA reductase in the crude extract. A 50 µL aliquot of the cell extract (29 mg/mL) was added to 10 µL 1M Tris-HCl (final concentration in assay 100 mM), 10 µL 10 mM malonyl CoA (final concentration in assay 1 mM), 5.5 µL 5.5 mM NADPH (final concentration in assay 0.3 mM), and 24.5 µL DI water in a 96 well UV transparent plate (Corning, NY). The enzyme activity was measured at 45°C using SpectraMAX Plus 96 well plate reader (Molecular devices Sunnyvale, CA). The activity of malonyl-CoA reductase was monitored by measuring the disappearance of NADPH at 340 nm wavelength. The crude extract exhibited malonyl-CoA reductase activity.

The 5 mL (total 145 mg) protein cell extract was diluted with 20 mL buffer A (20 mM ethanolamine (pH 9.0), 5 mM $MgCl_2$, 2 mM DTT). The chromatographic protein purification was conducted using a BioLogic protein purification system (BioRad Hercules, CA). The 25 mL of cell suspension was loaded onto a UNO Q-6 ion-exchange column that had been equilibrated with buffer A at a rate of 1 mL/minute. After sample loading, the column was washed with 2.5 times column volume of buffer A at a rate of 2 mL/minute. The proteins were eluted with a linear gradient of NaCl in buffer A from 0-0.33 M in 25 Column volume. During the entire chromatographic separation, three mL fractions were collected. The collection tubes contained 50 µL of Tris-HCl (pH 6.5) so that the pH of the eluted sample dropped to about pH 7. Major chromatographic peaks were detected in the region that corresponded to fractions 18 to 21 and 23 to 30. A 200 µL sample was taken from these fractions and concentrated in a microcentrifuge at 4°C using a Microcon YM-10 columns (Millipore Corp., Bedford, MA) as per manufacture's instructions. To each of the concentrated fraction, buffer A-Tris (100 mM Tris-HCl (pH 7.8), 5 mM $MgCl_2$, 2 mM DTT) was added to bring the total volume to 100 µL. Each of these fractions was tested for the malonyl-CoA reductase activity using the spectophotometric assay described above. The majority of specific malonyl CoA activity

was found in fractions 18 to 21. These fractions were pooled together, and the pooled sample was desalted using PD-10 column (Amersham Pharmacia Piscataway, NJ) as per manufacture's instructions.

The 10.5 mL of desalted protein extract was diluted with buffer A-Tris to a volume of 25 mL. This desalted diluted sample was applied to a 1 mL HiTrap Blue column (Amersham Pharmacia Piscataway, NJ) which was equilibrated with buffer A-Tris. The sample was loaded at a rate of 0.1 mL/minute. Unbound proteins were washed with 2.5 CV buffer A-Tris. The protein was eluted with 100 Mm Tris (pH 7.8), 5 mM $MgCl_2$, 2 mM DTT, 2mM NADPH, and 1 M NaCl. During this separation process, one mL fractions were collected. A 200 µL sample was drawn from fractions 49 to 54 and concentrated. Buffer A-Tris was added to each of the concentrated fractions to bring the total volume to 100 µL. Fractions were assayed for enzyme activity as described above. The highest specific activity was observed in fraction 51. The entire fraction 51 was concentrated as described above, and the concentrated sample was separated on an SDS-PAGE gel.

Electrophoresis was carried out using a Bio-Rad Protean II minigel system and pre-cast SDS-PAGE gels (4-15%), or a Protean II XI system and 16 cm x 20 cm x 1mm SDS-PAGE gels (10%) cast as per the manufacturer's protocol. The gels were run according to the manufacturer's instructions with a running buffer of 25 mM Tris-HCl (pH 8.3), 192 mM glycine, and 0.1% SDS.

A gel thickness of 1 mm was used to run samples from fraction 51. Protein from fraction 51 was loaded onto 10% SDS-PAGE (3 lanes, each containing 75 µg of total protein). The gels were stained briefly with Coomassie blue (Bio-Rad, Hercules, CA) and then destained to a clear background with a 10% acetic acid and 20% methanol solution. The staining revealed a band of about 130 to 140 KDa.

The protein band of about 130-140 KDa was excised with no excess unstained gel present. An equal area gel without protein was excised as a negative control. The gel slices were placed in uncolored microcentrifuge tubes, prewashed with 50% acetonitrile in HPLC-grade water, washed twice with 50% acetonitrile, and shipped on dry ice to Harvard Microchemistry Sequencing Facility, Cambridge, MA.

After *in-situ* enzymatic digestion of the polypeptide sample with trypsin, the resulting polypeptides were separated by micro-capillary reverse-phase HPLC. The HPLC was directly coupled to the nano-electrospray ionization source of a Finnigan LCQ quadrupole ion trap mass spectrometer (μLC/MS/MS). Individual sequence spectra (MS/MS) were acquired on-line at high sensitivity for the multiple polypeptides separated during the chromatographic run. The MS/MS spectra of the polypeptides were correlated with known sequences using the algorithm Sequest developed at the University of Washington (Eng *et al.*, *J. Am. Soc. Mass Spectrom.*, 5:976 (1994)) and programs developed at Harvard (Chittum *et al.*, *Biochemistry*, 37:10866 (1998)). The results were reviewed for consensus with known proteins and for manual confirmation of fidelity.

A similar purification procedure was used to obtain another sample (protein 1 sample) that was subjected to the same analysis that was used to evaluate the fraction 51 sample.

The polypeptide sequence results indicated that the polypeptides obtained from both the fraction 51 sample and the protein 1 sample had similarity to the six (764, 799, 859, 923, 1090, 1024) contigs sequenced from the C. *aurantiacus* genome and presented on the Joint Genome Institute's web site (http://www.jgi.doe.gov/). The 764 contig was the most prominent of the six with about 40 peptide sequences showing similarity. BLASTX analysis of each of these contigs on the GenBank web site (http://www.ncbi.nlm.nih.gov/BLAST/) indicated that the DNA sequence of the 764 contig (4201 bases) encoded for polypeptides that had a dehydrogenase/reductase type activity. Close inspection of the 764 contig, however, revealed that this contig did not have an appropriate ORF that would encode for a 130-140 KDa polypeptide.

BASLTX analysis also was conducted using the other five contigs. The results of this analysis were as follows. The 799 contig (3173 bases) appeared to encode polypeptides having phosphate and dehydrogenase type activities. The 859 contig (5865 bases) appeared to encode polypeptides having synthetase type activities. The 923 contig (5660 bases) appeared to encode polypeptides having elongation factor and synthetase type activities. The 1090 contig (15201 bases) appeared to encode polypeptides having dehydrogenase/reductase and cytochrome and sigma factor activities. The 1024 contig (12276 bases) appeared to encode polypeptides having dehydrogenase and decarboxylase

and synthetase type activities. Thus, the 859 and 923 contigs were eliminated from any further analysis.

The results from the BLASTX analysis also revealed that the dehydrogenase found in the 1024 contig was most likely an inositol monophosphate dehydrogenase. Thus, the 1024 contig was eliminated as a possible candidate that might encode for a polypeptide having malonyl-CoA reductase activity. The 799 contig also was eliminated since this contig is part of the OS17 polypeptide described above.

This narrowed down the search to 2 contigs, the 764 and 1090 contigs. Since the contigs were identified using the same protein sample and the dehydrogenase activities found in these contigs gave very similar BLASTX results, it was hypothesized that they are part of the same polypeptide. Additional evidence supporting this hypothesis was obtained from the discovery that the 764 and 1090 contigs are adjacent to each other in the *C. aurantiacus* genome as revealed by an analysis of scaffold data provided by the Joint Genome Institute. Sequence similarity and assembly analysis, however, revealed no overlapping sequence between these two contigs, possibly due to the presence of gaps in the genome sequencing.

The polypeptide sequences that belonged to the 764 and 1090 contigs were mapped. Based on this analysis, an appropriate coding frame and potential start and stop codons were identified. The following PCR primers were designed to PCR amplify a fragment that encoded for a polypeptide having malonyl-CoA reductase activity: PRO140F 5'-ATGGCGACGGGCGAGTCCATGAG-3', SEQ ID NO:153; PRO140R 5'-GGACACGAAGAACAGGGCGACAC-3', SEQ ID NO:154; and PRO140UP 5'-GAACTGTCTGGAGTAAGGCTGTC-3', SEQ ID NO:155. The PRO140F primer was designed based on the sequence of the 1090 contig and corresponds to the start of the potential start codon. The twelfth base was change from G to C to avoid primer-dimer formation. This change does not change the amino acid that was encoded by the codon. The PRO140R primer was designed based on sequence of the 764 contig and corresponds to a region located about 1 kB downstream from the potential stop codon. The PRO140UPF primer was designed based on sequence of the 1090 contig and corresponds to a region located about 300 bases upstream of potential start codon.

Genomic *C. aurantiacus* DNA was obtained. Briefly, *C. aurantiacus* was grown in 50 mL cultures for 3 to 4 days. Cells were pelleted and washed with 5 mL of a 10 mM Tris solution. The genomic DNA was then isolated using the gram positive bacteria protocol provided with Gentra Genomic "Puregene" DNA isolation kit (Gentra Systems, Minneapolis, MN). The cell pellet was resuspended in 1 mL Gentra Cell Suspension Solution to which 14.2 mg of lysozyme and 4 µL of 20 mg/mL proteinase K solution was added. The cell suspension was incubated at 37°C for 30 minutes. The precipitated genomic DNA was recovered by centrifugation at 3500g for 25 minutes and air-dried for 10 minutes. The genomic DNA was suspended in an appropriate amount of a 10 mM Tris solution and stored at 4°C.

Two PCR reactions were set-up using *C. aurantiacus* genomic DNA as template as follows:

| PCR Reaction #1 | | PCR program |
|---|---|---|
| 3.3 X *rTH* polymerase Buffer | 30 µL | 94°C 2 minutes |
| Mg(OAC) (25 mM) | 4 µL | 29 cycles of: |
| dNTP Mix (10 mM) | 3 µL | 94°C 30 seconds |
| PRO140F (100 µM) | 2 µL | 63°C 45 seconds |
| PRO140R (100 µM) | 2 µL | 68°C 4.5 minutes |
| Genomic DNA (100 ng/mL) | 1 µL | 68°C 7 minutes |
| *rTH* polymerase (2 U/µL) | 2 µL | 4°C Until further use |
| *pfu* polymerase (2.5 U/µL) | 0.25 µL | |
| DI water | 55.75 µL | |
| Total | 100 µL | |

| PCR Reaction #2 | | PCR program |
|---|---|---|
| 3.3 X *rTH* polymerase Buffer | 30 µL | 94°C 2 minutes |
| Mg(OAC) (25 mM) | 4 µL | 29 cycles of: |
| dNTP Mix (10 mM) | 3 µL | 94°C 30 seconds |
| PRO140UPF (100 µM) | 2 µL | 60°C 45 seconds |
| PRO140R (100 µM) | 2 µL | 68°C 4.5 minutes |

| | | | |
|---|---|---|---|
| Genomic DNA (100 ng/mL) | 1 µL | 68°C | 7 minutes |
| *rTH* polymerase (2 U/µL) | 2 µL | 4°C | Until further use |
| *pfu* polymerase 2.5 U/µL) | 0.25 µL | | |
| DI water | 55.75 µL | | |
| Total | 100 µL | | |

The products from both PCR reactions were separated on a 0.8% TAE gel. Both PCR reactions produced a product of 4.7 to 5 Kb in size. This approximately matched the expected size of a nucleic acid molecule that could encode a polypeptide having malonyl-CoA reductase activity.

Both PCR products were sequenced using sequencing primers (1090Fseq 5'-GATTCCGTATGTCACCCCTA-3', SEQ ID NO:156; and 764Rseq 5'-CAGGCGACTGGCAATCACAA-3', SEQ ID NO:157). The sequence analysis revealed a gap between the 764 and 1090 contigs. The nucleic acid sequence between the sequences from the764 and 1090 contigs was greater than 300 base pairs in length (Figure 51). In addition, the sequence analysis revealed an ORF of 3678 bases that showed similarities to dehydorgenase/reductase type enzymes (Figure 52). The amino acid sequence encoded by this ORF is 1225 amino acids in length (Figure 50). Also, BLASTP analysis of the amino acid sequence encoded by this ORF revealed two short chain dehydrogenase domains (adh type). These results are consistent with a polypeptide having malonyl-CoA reductase activity since such an enzyme involves two reduction steps for the conversion of malonyl CoA to 3-HP. Further, the computed MW of the polypeptide was determined to be about 134 KDa.

PCR was conducted using the PRO140F/PRO140R primer pair, *C. aurantiacus* genomic DNA, and the protocol described above as PCR reaction #1. After the PCR was completed, 0.25 U of *Taq* polymerase (Roche Molecular Biochemicals, Indianapolis, IN) was added to the PCR mix, which was then incubated at 72°C for 10 minutes. The PCR product was column purified using Qiagen PCR purification kit (Qiagen Inc., Valencia, CA). The purified PCR product was then TOPO cloned into expression vector pCRT7/CT as per manufacture's instructions (Invitrogen, Carlsbad, CA). TOP10 F' chemical competent cells were transformed with the TOPO ligation mix as per

manufacture's instructions (Invitrogen, Carlsbad, CA). The cells were recovered for half an hour, and the transformants were selected on LB/ampicillin (100 μg/mL) plates. Twenty single colonies were selected, and the plasmid DNA was isolated using Qiagen spin Mini prep kit (Qiagen Inc., Valencia, CA).

Each of these twenty clones were tested for correct orientation and right insert size by PCR. Briefly, plasmid DNA was used as a template, and the following two primers were used in the PCR amplification: PCRT7 5'-GAGACCACAACGGTTTCCCTCTA-3', SEQ ID NO:158; and PRO140R 5'-GGACACGAAGAACAGGGCGACAC-3', SEQ ID NO:159. The following PCR reaction mix and program was used:

| PCR Reaction | | PCR program | |
|---|---|---|---|
| 3.3 X *rTH* polymerase Buffer | 7.5 μL | 94°C | 2 minutes |
| Mg(OAC) (25 mM) | 1 μL | 25 cycles of: | |
| dNTP Mix (10 mM) | 0.5 μL | 94°C | 30 seconds |
| PCRT7 (100 μM) | 0.125 μL | 55°C | 45 seconds |
| PRO140R (100 μM) | 0.125 μL | 68°C | 4 minutes |
| Plasmid DNA | 0.5 μL | 68°C | 7 minutes |
| *rTH* polymerase (2 U/μL) | 0.5 μL | 4°C | Until further use |
| DI water | 14.75 μL | | |
| Total | 25 μL | | |

Out of twenty clone tested, only one clone exhibited the correct insert (Clone # P-10). Chemical competent cells of BL21(DE3)pLysS (Invitrogen, Carlsbad, CA) were transformed with 2 μL of the P-10 plasmid DNA as per the manufacture's instructions. The cells were recovered at 37°C for 30 minutes and were plated on LB ampicillin (100 μg/mL) and chloramphenicol (25 μg/mL).

A 20 mL culture of BL21(DE3)pLysS/P-10 and a 20 mL control culture of BL21(DE3)pLysS was incubated overnight. Using the overnight cultures as an inoculum, two 100 mL BL21(DE3)pLysS/P-10 clone cultures and two control strain cultures (BL21(DE3)pLysS) were started. All the cultures were induced with IPTG when they

reached an OD of about 0.5 at 600 nm. The control strain culture was induced with 10 μM IPTG or 100 μM IPTG, while one of the BL21(DE3)pLysS/P-10 clone cultures was induced with 10 μM IPTG and the other with 100 μM IPTG. The cultures were grown for 2.5 hours after induction. Aliquots were taken from each of the culture flasks before and after 2.5 hours of induction and separated using 4-15% SDS-PAGE to analyze polypeptide expression. In the induced BL21(DE3)pLysS/P-10 samples, a band corresponding to a polypeptide having a molecular weight of about 135 KDa was observed. This band was absent in the control strain samples and in samples taken before IPTG induction.

To assess malonyl-CoA reductase activity, BL21(DE3)pLysS/P-10 and BL21(DE3)pLysS cells were cultured and then harvested by centrifugation at 8,000 x g (Rotor JA 16.250, Beckman Coulter, Fullerton, CA). Once harvested, the cells were washed once with an equal volume of a 0.85% NaCl solution. The cell pellets were resuspended into 100 mM Tris-HCl buffer that was supplemented with 5 mM $Mg_2Cl$ and 2 mM DTT. The cells were disrupted by passing twice through a French Press Cell at 1,000 psi pressure (Gauge value). The cell debris was removed by centrifugation at 30,000 x g (Rotor JA 25.50, Beckman Coulter, Fullerton, CA). The cell extract was maintained at 4°C or on ice until further use.

Activity of malonyl-CoA reductase was measured at 37°C for both the control cells and the IPTG-induced cells. The activity of malonyl-CoA reductase was monitored by observing the disappearance of added NADPH as described above. No activity was found in the cell extract of the control strain, while the cell extract from the IPTG-induced BL21(DE3)pLysS/P-10 cells displayed malonyl-CoA reductase activity with a specific activity calculated to be about 0.0942 μmole/minute/mg of total protein.

Malonyl-CoA reductase activity also was measured by analyzing 3-HP formation from malonyl CoA using the following reaction conducted at 37°C:

| | Volume | Final conc. |
|---|---|---|
| Tris HCl (1M) | 10 μL | 100mM |
| Malonyl CoA (10mM) | 40 μL | 4 mM |
| NADPH (10 mM) | 30 μL | 3 mM |
| Cell extract | 20 μL | |

| | |
|---|---|
| Total | 100 µL |

The reaction was carried out at 37°C for 30 minutes. In the control reaction, a cell extract from BL21(DE3)pLysS was added to a final concentration of 322 mg total protein. In the experimental reaction mix, a cell extract from BL21(DE3)pLysS/P-10 was added to a final concentration of 226 mg of total protein. The reaction mixtures were frozen at -20°C until further analysis.

Chromatographic separation of the components in the reaction mixtures was performed using a HPX-87H (7.8x300mm) organic acid HPLC column (BioRad Laboratories , Hercules, CA). The column was maintained at 60°C. Mobile phase composition was HPLC grade water pH to 2.5 using triflouroacetic acid (TFA) and was delivered at a flow rate of 0.6 mL/minute.

Detection of 3-HP in the reaction samples was accomplished using a Waters/Micromass ZQ LC/MS instrument consisting of a Waters 2690 liquid chromatograph (Waters Corp., Milford, MA) with a Waters 996 Photo-diode Array (PDA) absorbance monitor placed in series between the chromatograph and the single quandrupole mass spectrometer. The ionization source was an Atmospheric Pressure Chemical Ionization (APCI) ionization source. All parameters of the APCI-MS system were optimized and selected based on the generation of the protonated molecular ion $([M+H])^+$ of 3-HP. The following parameters were used to detect 3-HP in the positive ion mode: Corona: 10 µA; Cone: 20V; Extractor: 2V; RF lens: 0.2V; Source temperature: 100°C; APCI Probe temperature: 300°C; Desolvation gas: 500L/hour; Cone gas: 50L/hour; Low mass resolution: 15; High mass resolution: 15; Ion energy: 1.0; Multiplier: 650. Data was collected in Selected Ion Reporting (SIR) mode set at m/z = 90.9.

Both the control reaction sample and the experimental reaction sample were probed for presence of 3-HP using the HPLC-mass spectroscopy technique. In the control samples, no 3-HP peak was observed, while the experimental sample exhibited a peak that matched the retention and the mass of 3-HP.

**Example 11 – Constructing recombinant cells that produce 3-HP**

A pathway to make 3-hydroxypropionate directly from glucose via acetyl CoA is presented in Figure 44. Most organisms such as *E. coli*, *Bacillus,* and yeast produce acetyl CoA from glucose via glycolysis and the action of pyruvate dehydrogenase. In order to divert the acetyl CoA generated from glucose, it is desirable to overexpress two genes, one encoding for acetyl CoA carboxylase and the other encoding malonyl-CoA reductase. As an example, these genes are expressed *in E. coli* through a T7 promoter using vectors pET30a and pFN476. The vector pET30a has a pBR *ori* and kanamycin resistance, while pFN476 has pSC101 *ori* and uses carbencillin resistance for selection. Because these two vectors have compatible *ori* and different markers they can be maintained in *E. coli* at the same time. Hence, the constructs used to engineer *E. coli* for direct production of 3-hydroxypropionate from glucose are pMSD8 (pFN476/accABCD) (Davis *et al.*, *J. Biol. Chem.*, 275:28593-28598, 2000) and pET30a/malonyl-CoA reductase or pET30a/*acc*1 and pFN476/malonyl-CoA reductase. The constructs are depicted in Figure 45.

To test the production of 3-hydroxypropionate from glucose, *E. coli* strain Tuner pLacI carrying plasmid pMSD8 (pFN476/accABCD) and pET30a/malonyl-CoA reductase or pET30a/*acc*1 and pFN476/malonyl-CoA reductase are grown in a B. Braun BIOSTAT B fermenter. A glass vessel fitted with a water jacket for heating is used to conduct this experiment. The fermenter working volume is1.5 L and is operated at 37°C. The fermenter is continuously supplied with oxygen by bubbling sterile air through it at a rate of 1 vvm. The agitation is cascaded to the dissolve oxygen concentration which is maintained at 40% DO. The pH of the liquid media is maintained at 7 using 2 N NaOH. The *E. coli* strain is grown in M9 media supplemented with 1% glucose, 1 μg/mL thiamine, 0.1% casamino acids, 10 μg/mL biotin, 50 μg/mL carbencillin, 50 μg/mL kanamycin, and 25 μg/mL chloramphenicol. The expression of the genes is induced when the cell density reached 0.5 OD(600nm) by adding 100 μM IPTG. After induction, samples of 2 mL volume are taken at 1, 2, 3, 4, and 8 hours. In addition, at 3 hours after induction, a 200 mL sample is taken to make a cell extract. The 2 mL samples are spun, and the supernatant is used to analyze products using LC/MS technique. The supernatant is stored at -20°C until further analysis.

The extract is prepared by spinning the 200 mL of cell suspension at 8000 g and washing the cell pellet with of 50 mL of 50 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 100 mM KCl, 2 mM DTT, and 5% glycerol. The cell suspension is spun again at 8000 g, and the pellet is resuspended into 5 mL of 50 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 100 mM KCl, 2 mM DTT, and 5% glycerol. The cells are disrupted by passing twice through a French Press at 1000 pisg. The cell debris is removed by centrifugation for 20 minutes at 30,000 g. All the operations are conducted at 4°C. To demonstrated *in vitro* formation of 3-hydroxypropionate using this recombinant cell extract, the following reaction of 200 µL is conducted at 37°C. The reaction mix is as follows: Tris HCl (pH 8.0; 100 mM), ATP (1 mM), MgCl2 (5 mM), KCl (100 mM), DTT (5 mM), NaHCO3 (40 mM), NADPH (0.5 mM), acetyl CoA (0.5 mM), and cell extract (0.2 mg). The reaction is stopped after 15 minutes by adding 1 volume of 10% trifluroacetic acid (TFA). The products of this reaction are detected using an LC/MS technique.

The detection and analysis for the presence of 3-hydroxypropionate in the supernatant and the *in vitro* reaction mixture is carried out using a Waters/Micromass ZQ LC/MS instrument. This instrument consists of a Waters 2690 liquid chromatograph with a Waters 2410 refractive index detector placed in series between the chromatograph and the single quadropole mass spectrometer. LC separations are made using a Bio-Rad Aminex 87-H ion-exchange column at 45°C. Sugars, alcohol, and organic acid products are eluted with 0.015% TFA buffer. For elution, the buffer is passed at a flow rate of 0.6 mL/minute. For detection and quantification of 3-hydroxypropionate, a sample obtained from TCI, America (Portland, OR) is used as a standard.

**Example 12 Cloning of propionyl-CoA transferase, lactyl-CoA dehydratase (LDH), and a hydratase (OS19) for Expression in *Saccharomyces cerevisiae***

The pESC Yeast Epitope Tagging Vector System (Stratagene, La Jolla, CA) was used in cloning the genes involved in 3-hydroxypropionic acid production via lactic acid. The pESC vectors each contain GAL1 and GAL10 promoters in opposing directions, allowing the expression of two genes from each vector. The GAL1 and GAL10 promoters are repressed by glucose and induced by galactose. Each of the four available pESC vectors has a different yeast-selectable marker (HIS3, TRP1, LEU2, URA3) so

multiple plasmids can be maintained in a single strain. Each cloning region has a polylinker site for gene insertion, a transcription terminator, and an epitope coding sequence for C-terminal or N-terminal epitope tagging of expressed proteins. The pESC vectors also have a ColE1 origin of replication and an ampicillin resistance gene to allow replication and selection in *E. coli*. The following vector/promoter/nucleic acid combinations were constructed:

| Vector | Promoter | Polypeptide | Source of nucleic acid |
|---|---|---|---|
| pESC-Trp | GAL1 | OS19 hydratase | *Chloroflexus aurantiacus* |
| | GAL10 | E1 | *Megasphaera elsdenii* |
| pESC-Leu | GAL1 | E2α | *Megasphaera elsdenii* |
| | GAL10 | E2β | *Megasphaera elsdenii* |
| pESC-His | GAL1 | D-LDH | *Escherishia coli* |
| | GAL10 | PCT | *Megasphaera elsdenii* |

The primers used were as follows:

OS19APAF: 5'-ATAGGGCCCAGGAGATCAAACCATGGGTGAAGAGTCT-CTGGTTC-3' (SEQ ID NO:164)

OS19SALR: 5'-CCTCTGCTACAGTCGACACAACGACCACTGAAGTTG-GGAG-3'(SEQ ID NO:165)

OS19KPNR: 5'-AGTCTGCTATCGGTACCTCAACGACCACTGAAGTTG-GGAG-3'(SEQ ID NO:166)

EINOTF: 5'-ATAGCGGCCGCATAATGGATACTCTCGGAATCGACG-TTGG-3'(SEQ ID NO:167)

EICLAR: 5'-CCCCATCGATACATATTTCTTGATTTTATCATAAGCA-ATC-3'(SEQ ID NO:168)

EIIαAPAF: 5'-CCAGGGCCCATAATGGGTGAAGAAAAAACAGTAGA-TATTG-3'(SEQ ID NO:169)

EIIαSALR: 5'-GGTAGACTTGTCGACGTAGTGGTTTCCTCCTTCATT-GG-3'(SEQ ID NO:170)

EIIβNOTF: 5'-ATAGCGGCCGCATAATGGGTCAGATCGACGAACTTA-

TCAG-3'(SEQ ID NO:171)

EIIβSPER: 5'-AGGTTCAACTAGTTCGTAGAGGATTTCCGAGAAAGC-CTG-3'(SEQ ID NO:172)

LDHAPAF: 5'-CTAGGGCCCATAATGGAACTCGCCGTTTATAG-CAC-3'(SEQ ID NO:173)

LDHXHOR: 5'-ACTTCTCGAGTTAAACCAGTTCGTTCGGGCA-GGT-3'(SEQ ID NO:174)

PCTSPEF: 5'-GGGACTAGTATAATGGGAAAAGTAGAAATCAT-TACAG-3'(SEQ ID NO:175)

PCTPACR: 5'-CGGCTTAATTAACAGCAGAGATTTATTTTTTCA-GTCC-3'(SEQ ID NO:176)

All restriction enzymes were obtained from New England Biolabs, Beverly, MA. All plasmid DNA preparations were done using QIAprep Spin Miniprep Kits, and all gel purifications were done using QIAquick Gel Extraction Kits (Qiagen, Valencia, CA).

A. Construction of the pESC-Trp/OS19 hydratase vector

Two constructs in pESC-Trp were made for the OS19 nucleic acid from *C. aurantiacus*. One of these constructs utilized the *Apa* I and *Sal* I restriction sites of the GAL1 multiple cloning site and was designed to include the c-myc epitope. The second construct utilized the *Apa* I and *Kpn* I sites and thus did not include the c-myc epitope sequence.

Six μg of pESC-Trp vector DNA was digested with the restriction enzyme *Apa* I and the digest was purified using a QIAquick PCR Purification Column. Three μg of the *Apa* I-digested vector DNA was then digested with the restriction enzyme *Kpn* I, and 3 μg was digested with *Sal* I. The double-digested vector DNAs were separated on a 1% TAE-agarose gel, purified, dephosphorylated with shrimp alkaline phosphatase (Roche Biochemical Products, Indianapolis, IN), and purified with a QIAquick PCR Purification Column.

The nucleic acid encoding the *Chloroflexus aurantiacus* polypeptide having hydratase activity (OS19) was amplified from genomic DNA using the PCR primer pair OS19APAF and OS19SALR and the primer pair OS19APAF and OS19KPNR.

OS19APAF was designed to introduce an *Apa* I restriction site and a translation initiation site (ACCATGG) at the beginning of the amplified fragment. The OS19KPNR primer was designed to introduce a *Kpn* I restriction site at the end of the amplified fragment and to contain the translational stop codon for the hydratase gene. OS19SALR introduces a *Sal* I site at the end of the amplified fragment and has an altered stop codon so that translation continues in-frame through the vector c-myc epitope. The PCR mix contained the following: 1X Expand PCR buffer, 100 ng *C. aurantiacus* genomic DNA, 0.2 μM of each primer, 0.2 mM each dNTP, and 5.25 units of Expand DNA Polymerase (Roche) in a final volume of 100 μL. The PCR reaction was performed in an MJ Research PTC100 under the following conditions: an initial denaturation at 94°C for 1 minute; 8 cycles of 94°C for 30 seconds, 57°C for 1 minute, and 72°C for 2.25 minutes; 24 cycles of 94°C for 30 seconds, 62°C for 1 minute, and 72°C for 2.25 minutes; and a final extension for 7 minutes at 72°C. The amplification product was then separated by gel electrophoresis using a 1% TAE-agarose gel. A 0.8 Kb fragment was excised from the gel and purified for each primer pair. The purified fragments were digested with *Kpn* I or *Sal* I restriction enzyme, purified with a QIAquick PCR Purification Column, digested with *Apa* I restriction enzyme, purified again with a QIAquick PCR Purification Column, and quantified on a minigel.

50-60 ng of the digested PCR product containing the nucleic acid encoding the *C. aurantiacus* polypeptide having hydratase activity (OS19) and 50 ng of the prepared pESC-Trp vector were ligated using T4 DNA ligase at 16°C for 16 hours. One μL of the ligation reaction was used to electroporate 40 μL of *E. coli* Electromax™ DH10B™ cells. The electroporated cells were plated onto LB plates containing 100 μg/mL of carbenicillin (LBC). Individual colonies were screened using colony PCR with the appropriate PCR primers. Individual colonies were suspended in about 25 μL of 10 mM Tris, and 2 μL of the suspension was plated on LBC media. The remnant suspension was heated for 10 minutes at 95°C to break open the bacterial cells, and 2 μL of the heated cells was used in a 25 μL PCR reaction. The PCR mix contained the following: 1X Taq buffer, 0.2 μM each primer, 0.2 mM each dNTP, and 1 unit of Taq DNA polymerase per reaction. The PCR program used was the same as described above for amplification of the nucleic acid from genomic DNA.

Plasmid DNA was isolated from cultures of colonies having the desired insert and was sequenced to confirm the lack of nucleotide errors from PCR. A construct with a confirmed sequence was transformed into *S. cerevisiae* strain YPH500 using a Frozen-EZ Yeast Transformation II™ Kit (Zymo Research, Orange, CA). Transformation reactions were plated on SC-Trp media (see Stratagene pESC Vector Instruction Manual for media recipes). Individual yeast colonies were screened for the presence of the OS19 nucleic acid by colony PCR. Colonies were suspended in 20 µL of Y-Lysis Buffer (Zymo Research) containing 5 units of zymolase and heated at 37°C for 10 minutes. Three µL of this suspension was then used in a 25 µL PCR reaction using the PCR reaction mixture and program described for the colony screen of the *E. coli* transformants. The pESC-Trp vector was also transformed into YPH500 for use as a hydratase assay control and transformants were screened by PCR using GAL1 and GAL10 primers.

B. Construction of the pESC-Trp/OS19/EI hydratase vector

Plasmid DNA of a pESC-Trp/OS19 construct (*Apa* I-*Sal* I sites) with confirmed sequence and positive assay results was used for insertion of the nucleic acid for the *M. elsdenii* E1 activator polypeptide downstream of the GAL10 promoter. Three µg of plasmid DNA was digested with the restriction enzyme *Cla* I, and the digest was purified using a QIAquick PCR Purification Column. The vector DNA was then digested with the restriction enzyme *Not* I, and the digest was inactivated by heating to 65°C for 20 minutes. The double-digested vector DNA was dephosphorylated with shrimp alkaline phosphatase (Roche), separated on a 1% TAE-agarose gel, and gel purified.

The nucleic acid encoding the *M. elsdenii* E1 activator polypeptide was amplified from genomic DNA using the PCR primer pair EINOTF and EICLAR. EINOTF was designed to introduce a *Not* I restriction site and a translation initiation site at the beginning of the amplified fragment. The EICLAR primer was designed to introduce a *Cla* I restriction site at the end of the amplified fragment and to contain an altered translational stop codon to allow in-frame translation of the FLAG epitope. The PCR mix contained the following: 1X Expand PCR buffer, 100 ng *M. elsdenii* genomic DNA, 0.2 µM of each primer, 0.2 mM each dNTP, and 5.25 units of Expand DNA Polymerase in a final volume of 100 µL. The PCR reaction was performed in an MJ Research PTC100

under the following conditions: an initial denaturation at 94°C for 1 minute; 8 cycles of 94°C for 30 seconds, 55°C for 45 seconds, and 72°C for 3 minutes; 24 cycles of 94°C for 30 seconds, 62°C for 45 seconds, and 72°C for 3 minutes; and a final extension for 7 minutes at 72°C. The amplification product was then separated by gel electrophoresis using a 1% TAE-agarose gel, and a 0.8 Kb fragment was excised and purified. The purified fragment was digested with *Cla* I restriction enzyme, purified with a QIAquick PCR Purification Column, digested with *Not* I restriction enzyme, purified again with a QIAquick PCR Purification Column, and quantified on a minigel.

60 ng of the digested PCR product containing the nucleic acid for the *M. elsdenii* E1 activator polypeptide and 70 ng of the prepared pESC-Trp/OS19 hydratase vector were ligated using T4 DNA ligase at 16°C for 16 hours. One µL of the ligation reaction was used to electroporate 40 µL of *E. coli* Electromax™ DH10B™ cells. The electroporated cells were plated onto LBC media. Individual colonies were screened using colony PCR with the EINOTF and EICLAR primers. Individual colonies were suspended in about 25 µL of 10 mM Tris, and 2 µL of the suspension was plated on LBC media. The remnant suspension was heated for 10 minutes at 95°C to break open the bacterial cells, and 2 µL of the heated cells used in a 25 µL PCR reaction. The PCR mix contained the following: 1X Taq buffer, 0.2 µM each primer, 0.2 mM each dNTP, and 1 unit of Taq DNA polymerase per reaction. The PCR program used was the same as described above for amplification of the gene from genomic DNA. Plasmid DNA was isolated from cultures of colonies having the desired insert and was sequenced to confirm the lack of nucleotide errors from PCR.

C. Construction of the pESC-Leu/EIIα/EIIβ vector

Three µg of DNA of the vector pESC-Leu was digested with the restriction enzyme *Apa* I, and the digest was purified using a QIAquick PCR Purification Column. The vector DNA was then digested with the restriction enzyme *Sal* I, and the digest was inactivated by heating to 65°C for 20 minutes. The double-digested vector DNA was dephosphorylated with shrimp alkaline phosphatase (Roche), separated on a 1% TAE-agarose gel, and gel purified.

The nucleic acid encoding the *M. elsdenii* E2α polypeptide was amplified from genomic DNA using the PCR primer pair EIIαAPAF and EIIaαSALR. EIIαAPAF was designed to introduce an *Apa* I restriction site and a translation initiation site at the beginning of the amplified fragment. The EIIαSALR primer was designed to introduce a *Sal* I restriction site at the end of the amplified fragment and to contain an altered translational stop codon to allow in-frame translation of the c-myc epitope. The PCR mix contained the following: 1X Expand PCR buffer, 100 ng *M. elsdenii* genomic DNA, 0.2 μM of each primer, 0.2 mM each dNTP, and 5.25 units of Expand DNA Polymerase in a final volume of 100 μL. The PCR reaction was performed in an MJ Research PTC100 under the following conditions: an initial denaturation at 94°C for 1 minute; 8 cycles of 94°C for 30 seconds, 55°C for 1 minute, and 72°C for 3 minutes; 24 cycles of 94°C for 30 seconds, 62°C for 1 minute, and 72°C for 3 minutes; and a final extension for 7 minutes at 72°C. The amplification product was then separated by gel electrophoresis using a 1% TAE-agarose gel, and a 1.3 Kb fragment was excised and purified. The purified fragment was digested with *Apa* I restriction enzyme, purified with a QIAquick PCR Purification Column, digested with *Sal* I restriction enzyme, purified again with a QIAquick PCR Purification Column, and quantified on a minigel.

80 ng of the digested PCR product containing the nucleic acid encoding the *M. elsdenii* E2α polypeptide and 80 ng of the prepared pESC-Leu vector were ligated using T4 DNA ligase at 16°C for 16 hours. One μL of the ligation reaction was used to electroporate 40 μL of *E. coli* Electromax™ DH10B™ cells. The electroporated cells were plated onto LBC media. Individual colonies were screened using colony PCR with the EIIαAPAF and EIIαSALR primers. Individual colonies were suspended in about 25 μl of 10 mM Tris, and 2 μL of the suspension was plated on LBC media. The remnant suspension was heated for 10 minutes at 95°C to break open the bacterial cells, and 2 μL of the heated cells used in a 25 μL PCR reaction. The PCR mix contained the following: 1X Taq buffer, 0.2 μM each primer, 0.2 mM each dNTP, and 1 unit of Taq DNA polymerase per reaction. The PCR program used was the same as described above for amplification of the gene from genomic DNA. Plasmid DNA was isolated from cultures of colonies having the desired insert and was sequenced to confirm the lack of nucleotide errors from PCR.

Plasmid DNA of a pESC-Leu/EIIα vector with confirmed sequence was used for insertion of the nucleic acid encoding the *M. elsdenii* E2β polypeptide. Three μg of plasmid DNA was digested with the restriction enzyme *Spe* I, and the digest was purified using a QIAquick PCR Purification Column. The vector DNA was then digested with the restriction enzyme *Not* I and gel purified from a 1% TAE-agarose gel. The double-digested vector DNA was then dephosphorylated with shrimp alkaline phosphatase (Roche) and purified with a QIAquick PCR Purification Column.

The nucleic acid encoding the *M. elsdenii* E2β polypeptide was amplified from genomic DNA using the PCR primer pair EIIβNOTF and EIIβSPER. The EIIβNOTF primer was designed to introduce a *Not* I restriction site and a translation initiation site at the beginning of the amplified fragment. The EIIβSPER primer was designed to introduce an *Spe* I restriction site at the end of the amplified fragment and to contain an altered translational stop codon to allow for in-frame translation of the FLAG epitope. The PCR mix contained the following: 1X Expand PCR buffer, 100 ng *M. elsdenii* genomic DNA, 0.2 μM of each primer, 0.2 mM each dNTP, and 5.25 units of Expand DNA Polymerase in a final volume of 100 μL. The PCR reaction was performed in an MJ Research PTC100 under the following conditions: an initial denaturation at 94°C for 1 minute; 8 cycles of 94°C for 30 seconds, 55°C for 45 seconds, and 72°C for 3 minutes; 24 cycles of 94°C for 30 seconds, 62°C for 45 seconds, and 72°C for 3 minutes; and a final extension for 7 minutes at 72°C. The amplification product was separated by gel electrophoresis using a 1% TAE-agarose gel, and a 1.1 Kb fragment was excised and purified. The purified fragment was digested with *Spe* I restriction enzyme, purified with a QIAquick PCR Purification Column, digested with *Not* I restriction enzyme, purified again with a QIAquick PCR Purification Column, and quantified on a minigel.

38 ng of the digested PCR product containing the nucleic acid encoding the *M. elsdenii* E2β polypeptide and 50 ng of the prepared pESC-Leu/EIIα vector were ligated using T4 DNA ligase at 16°C for 16 hours. One μL of the ligation reaction was used to electroporate 40 μL of *E. coli* Electromax™ DH10B™ cells. The electroporated cells were plated onto LBC plates. Individual colonies were screened using colony PCR with the EIIβNOTF and EIIβSPER primers. Individual colonies were suspended in about 25 μL of 10 mM Tris, and 2 μL of the suspension was plated on LBC media. The remnant

suspension was heated for 10 minutes at 95°C to break open the bacterial cells, and 2 μL of the heated cells was used in a 25 μL PCR reaction. The PCR mix contained the following: 1X Taq buffer, 0.2 μM each primer, 0.2 mM each dNTP, and 1 unit of Taq DNA polymerase per reaction. The PCR program used was the same as described above for amplification of the gene from genomic DNA.

Plasmid DNA was isolated from cultures of colonies having the desired insert and was sequenced to confirm the lack of nucleotide errors from PCR. A pESC-Leu/EIIα /EIIβ construct with a confirmed sequence was co-transformed along with the pESC-Trp/OS19/EI vector into *S. cerevisiae* strain YPH500 using a Frozen-EZ Yeast Transformation II™ Kit (Zymo Research, Orange, CA). Transformation reactions were plated on SC-Trp-Leu media. Individual yeast colonies were screened for the presence of the OS19, E1, E2α, and E2β nucleic acid by colony PCR. Colonies were suspended in 20 μL of Y-Lysis Buffer (Zymo Research) containing 5 units of zymolase and heated at 37°C for 10 minutes. Three μL of this suspension was then used in a 25 μL PCR reaction using the PCR reaction mixtures and programs described for the colony screens of the *E. coli* transformants. The pESC-Trp/OS19 and pESC-Leu vectors were also co-transformed intoYPH500 for use as a lactyl-CoA dehydratase assay control. These transformants were colony screened using the GAL1 and GAL10 primers (Instruction manual, pESC Yeast Epitope Tagging Vectors, Stratagene).

D. Construction of the pESC-His/D-LDH/PCT vector

Three μg of DNA of the vector pESC-His was digested with the restriction enzyme *Xho* I, and the digest was purified using a QIAquick PCR Purification Column. The vector DNA was then digested with the restriction enzyme *Apa* I and gel purified from a 1% TAE-agarose gel. The double-digested vector DNA was dephosphorylated with shrimp alkaline phosphatase (Roche) and purified using a QIAquick PCR Purification Column.

The *E. coli* D-LDH gene was amplified from genomic DNA of strain DH10B using the PCR primer pair LDHAPAF and LDHXHOR. LDHAPAF was designed to introduce an *Apa* I restriction site and a translation initiation site at the beginning of the amplified fragment. The LDHXHOR primer was designed to introduce an *Xho* I

restriction site at the end of the amplified fragment and to contain the translational stop codon for the D-LDH gene. The PCR mix contained the following: 1X Expand PCR buffer, 100 ng *E. coli* genomic DNA, 0.2 μM of each primer, 0.2 mM each dNTP, and 5.25 units of Expand DNA Polymerase in a final volume of 100 μL. The PCR reaction was performed in an MJ Research PTC100 under the following conditions: an initial denaturation at 94°C for 1 minute; 8 cycles of 94°C for 30 seconds, 59°C for 45 seconds, and 72°C for 2 minutes; 24 cycles of 94°C for 30 seconds, 64°C for 45 seconds, and 72°C for 2 minutes; and a final extension for 7 minutes at 72°C. The amplification product was separated by gel electrophoresis using a 1% TAE-agarose gel, and a 1.0 Kb fragment was excised and purified. The purified fragment was digested with *Apa* I restriction enzyme, purified with a QIAquick PCR Purification Column, digested with *Xho* I restriction enzyme, purified again with a QIAquick PCR Purification Column, and quantified on a minigel.

80 ng of the digested PCR product containing the *E. coli* D-LDH gene and 80 ng of the prepared pESC-His vector were ligated using T4 DNA ligase at 16°C for 16 hours. One μL of the ligation reaction was used to electroporate 40 μL of *E. coli* Electromax™ DH10B™ cells. The electroporated cells were plated onto LBC media. Individual colonies were screened using colony PCR with the LDHAPAF and LDHXHOR primers. Individual colonies were suspended in about 25 μL of 10 mM Tris, and 2 μL of the suspension was plated on LBC media. The remnant suspension was heated for 10 minutes at 95°C to break open the bacterial cells, and 2 μL of the heated cells used in a 25 μL PCR reaction. The PCR mix contained the following: 1X Taq buffer, 0.2 μM each primer, 0.2 mM each dNTP, and 1 unit of Taq DNA polymerase per reaction. The PCR program used was the same as described above for amplification of the gene from genomic DNA. Plasmid DNA was isolated from cultures of colonies having the desired insert and was sequenced to confirm the lack of nucleotide errors from PCR.

Plasmid DNA of a pESC-His/D-LDH construct with a confirmed sequence was used for insertion of the nucleic acid encoding the *M. elsdenii* PCT polypeptide. Three μg of plasmid DNA was digested with the restriction enzyme *Pac* I, and the digest was purified using a QIAquick PCR Purification Column. The vector DNA was then digested with the restriction enzyme *Spe* I and gel purified from a 1% TAE-agarose gel. The

double-digested vector DNA was dephosphorylated with shrimp alkaline phosphatase (Roche) and purified with a QIAquick PCR Purification Column.

The nucleic acid encoding the *M. elsdenii* PCT polypeptide was amplified from genomic DNA using the PCR primer pair PCTSPEF and PCTPACR. PCTSPEF was designed to introduce an *Spe* I restriction site and a translation initiation site at the beginning of the amplified fragment. The PCTPACR primer was designed to introduce a *Pac* I restriction site at the end of the amplified fragment and to contain the translational stop codon for the PCT gene. The PCR mix contained the following: 1X Expand PCR buffer, 100 ng M. elsdenii genomic DNA, 0.2 μM of each primer, 0.2 mM each dNTP, and 5.25 units of Expand DNA Polymerase in a final volume of 100 μL. The PCR reaction was performed in an MJ Research PTC100 under the following conditions: an initial denaturation at 94°C for 1 minute; 8 cycles of 94°C for 30 seconds, 56°C for 45 seconds, and 72°C for 2.5 minutes; 24 cycles of 94°C for 30 seconds, 64°C for 45 seconds, and 72°C for 2.5 minutes; and a final extension for 7 minutes at 72°C. The amplification product was separated by gel electrophoresis using a 1% TAE-agarose gel, and a 1.55 Kb fragment was excised and purified. The purified fragment was digested with *Pac* I restriction enzyme, purified with a QIAquick PCR Purification Column, digested with *Spe* I restriction enzyme, purified again with a QIAquick PCR Purification Column, and quantified on a minigel.

95 ng of the digested PCR product containing the nucleic acid encoding the *M. elsdenii* PCT polypeptide and 75 ng of the prepared pESC-His/D-LDH vector were ligated using T4 DNA ligase at 16°C for 16 hours. One μL of the ligation reaction was used to electroporate 40 μL of *E. coli* Electromax™ DH10B™ cells. The electroporated cells were plated onto LBC plates. Individual colonies were screened using colony PCR with the PCTSPEF and PCTPACR primers. Individual colonies were suspended in about 25 μL of 10 mM Tris, and 2 μL of the suspension was plated on LBC media. The remnant suspension was heated for 10 minutes at 95°C to break open the bacterial cells, and 2 μL of the heated cells used in a 25 μL PCR reaction. The PCR mix contained the following: 1X Taq buffer, 0.2 μM each primer, 0.2 mM each dNTP, and 1 unit of Taq DNA polymerase per reaction. The PCR program used was the same as described above for amplification of the gene from genomic DNA.

Plasmid DNA was isolated from cultures of colonies having the desired insert and was sequenced to confirm the lack of nucleotide errors from PCR. A construct with a confirmed sequence was transformed into *S. cerevisiae* strain YPH500 using a Frozen-EZ Yeast Transformation II™ Kit (Zymo Research, Orange, CA). Transformation reactions were plated on SC-His media. Individual yeast colonies were screened for the presence of the D-LDH and PCT genes by colony PCR. Colonies were suspended in 20 µL of Y-Lysis Buffer (Zymo Research) containing 5 units of zymolase and heated at 37°C for 10 minutes. Three µL of this suspension was then used in a 25 µL PCR reaction using the PCR reaction mixture and program described for the colony screen of the *E. coli* transformants. The pESC-His vector was also transformed into YPH500 for use as an assay control, and transformants were screened by PCR using GAL1 and GAL10 primers.

## Example 13 - Expression of Enzymes in *S. cerevisiae*

A. Hydratase Activity in Transformed Yeast

Individual colonies carrying the pESC-Trp/OS19 construct or the pESC-Trp vector (negative control) were used to inoculate 5 mL cultures of SC-Trp media containing 2% glucose. These cultures were grown for 16 hours at 30°C and used to inoculate 35 mL of the same media. The subcultures were grown for 7 hours at 30°C, and their $OD_{600}$s were determined. A volume of cells giving an OD x volume equal to 40 was pelleted, washed with SC-Trp media with no carbon source, and repelleted. The cells were suspended in 5 mL of SC-Trp media containing 2% galactose and used to inoculate a total volume of 100 mL of this media. Cultures were grown for 17.5 hours at 30°C and 250 rpm. Cells were then pelleted, rinsed in 0.85% NaCl, and repelleted. Cell pellets (70 mg) were suspended in 140 µL of 50 mM TrisHCl, pH 7.5, and an equal volume (pellet plus buffer) of pre-rinsed glass beads (Sigma, 150-212 microns) was added. This mixture was vortexed for 30 seconds and placed on ice for 1 minute, and the vortexing/cooling cycle was repeated 8 additional times. The cells were then centrifuged for 6 minutes at 5,000*g*, and the supernatant was removed to a fresh tube. The beads/pellet were washed twice with 250 µL of buffer, centrifuged, and the supernatants joined with the first supernatant.

An *E. coli* strain carrying the pETBlue-1/OS19 construct, described previously, was used as a positive control for hydratase assays. A culture of this strain was grown to saturation overnight and diluted 1:20 the following morning in fresh LBC media. The culture was grown at 37°C and 250 rpm to an $OD_{600}$ of 0.6, at which point it was induced with IPTG at a final concentration of 1 mM. The culture was incubated for an additional two hours at 37°C and 250 rpm. Cells were pelleted, washed with 0.85 % NaCl, and repelleted. Cells were disrupted using BugBuster™ Protein Extraction Reagent and Benzonase® (Novagen) as per manufacturer's instructions with a 20 minute incubation at room temperature. After centrifugation at 16,000*g* and 4°C, the supernatant was transferred to a new tube and used in the activity assay.

Total protein content of cell extracts from *S. cerevisiae* described above were quantified using a microplate Bio-Rad Protein Assay (Bio-Rad, Hercules, CA). The OS19 constructs (both *Apa* I-*Sal* I and *Apa* I-*Kpn* I sites) in YPH500, the pESC-Trp negative control in YPH500, and the pETBlue-1/OS19 construct in *E. coli* were tested for their ability to convert acrylyl-CoA to 3-hydroxypropionyl-CoA. The assay was conducted as previously described for the pETBlue-1/OS19 constructs in the *E. coli* Tuner strain. When cell extract of the negative control strain was added to the reaction mixture containing acrylyl-CoA, one dominant peak of MW 823 was exhibited. This peak corresponds to acrylyl-CoA and indicates that acrylyl-CoA was not converted to any other product. When cell extract of the strain carrying a pESC-Trp/OS19 construct (either *Apa* I-*Sal* I or *Apa* I-*Kpn* I sites) was added to the reaction mix, the dominant peak shifted to MW 841, which corresponds to 3-hydroxypropionyl-CoA. The reaction mix from the *E. coli* control also showed the MW 841 peak. A time course study was conducted for the pESC-Trp/OS19(*Apa* I-*Sal* I) construct, which measured the appearance of the MW 841 and MW 823 peaks after 0, 1, 3, 7, 15, 30, and 60 minutes of reaction time. An increase in the 3-hydroxypropionyl-CoA peak was seen over time with the cell extracts from both this construct and the *E. coli* control, whereas cell extract from the YPH500 strain with vector only showed a dominant acrylyl-CoA peak.

B. Propionyl CoA-Transferase Activity in Transformed Yeast

Individual colonies of *S. cerevisiae* strain YPH500 carrying the pESC-His/D-LDH or pESC-His/D-LDH/PCT construct or the pESC-His vector with no insert (negative control) were used to inoculate 5 mL cultures of SC-His media containing 2% glucose. These cultures were grown for 16 hours at 30°C and 250 rpm and were then used to inoculate 35 mL of the same media. The subcultures were grown for 7 hours at 30°C, and their $OD_{600}$s were determined. For each strain, a volume of cells giving an OD x volume equal to 40 was pelleted, washed with SC-His media with no carbon source, and repelleted. The cells were suspended in 5 mL of SC-His media containing 2% galactose and used to inoculate a total volume of 100 mL of this media. Cultures were grown for 16.75 hours at 30°C and 250 rpm. Cells were then pelleted, rinsed in 0.85% NaCl, and repelleted. Cell pellets (70 mg) were suspended in 140 µL of 100 mM potassium phosphate buffer, pH 7.5, and an equal volume (pellet plus buffer) of pre-rinsed glass beads (Sigma, 150-212 microns) was added. This mixture was vortexed for 30 seconds and placed on ice for 1 minute, and the vortexing/cooling cycle was repeated 8 additional times. The cells were then centrifuged for 6 minutes at 5,000g, and the supernatant was removed to a fresh tube. The beads/pellet were washed twice with 250 µL of buffer and centrifuged, and the supernatants joined with the first supernatant.

An *E. coli* strain carrying the pETBlue-1/PCT construct, described previously, was used as a positive control for propionyl CoA transferase assays. A culture of this strain was grown to saturation overnight and diluted 1:20 the following morning in fresh LB media containing 100 µg/mL of carbenicillin. The culture was grown at 37°C and 250 rpm to an $OD_{600}$ of 0.6, at which point it was induced with IPTG at a final concentration of 1 mM. The culture was incubated for an additional two hours at 37°C and 250 rpm. Cells were pelleted, washed with 0.85 % NaCl, and repelleted. Cells were disrupted using BugBuster™ Protein Extraction Reagent and Benzonase® (Novagen) as per manufacturer's instructions with a 20 minute incubation at room temperature. After centrifugation at 16,000*g* and 4°C, the supernatant was transferred to a new tube and used in the activity assay.

Total protein content of cell extracts was quantified using a microplate Bio-Rad Protein Assay (Bio-Rad, Hercules, CA). The D-LDH and D-LDH/PCT constructs in *S.*

*cerevisiae* strain YPH500, the pESC-His negative control in YPH500, and the pETBlue-1/PCT construct in *E. coli* were tested for their ability to catalyze the conversion of propionyl-CoA and acetate to acetyl-CoA and propionate. The assay mixture used was that previously described for the pETBlue-1/PCT constructs in the *E. coli* Tuner strain.

When 1 μg of total cell extract protein of the negative control strain or the YPH500/pESC-His/D-LDH strain was added to the reaction mixture, no increase in absorbance (0.00 to 0.00) was seen over 11 minutes. Increases in absorbance from 0.00 to 0.04 and from 0.00 to 0.06 were seen, respectively, with 1 μg of cell extract protein from the YPH500/pESC-His/D-LDH/PCT strain and the *E. coli*/PCT strain. With 2 mg of total cell extract protein, the negative control strain and the YPH500/pESC-His/D-LDH strain showed an increase in absorbance from 0.00 to 0.01 over 11 minutes, whereas increases from 0.00 to 0.10 and 0.00 to 0.08 were seen, respectively, with the YPH500/pESC-His/ D-LDH /PCT strain and the *E. coli*/PCT strain.

C. Lactyl-CoA Dehydratase Activity in Transformed Yeast

Individual colonies of *S. cerevisiae* strain YPH500 carrying the pESC-His/D-LDH or pESC-His/D-LDH/PCT construct or the pESC-His vector with no insert (negative control) were used to inoculate 5 mL cultures of SC-His media containing 4% glucose. These cultures were grown for 23 hours at 30°C and used to inoculate 35 mL of SC-His media containing 2 % raffinose. The subcultures were grown for 8 hours at 30°C, and their $OD_{600}$s were determined. For each strain, a volume of cells giving an OD x volume equal to 40 was pelleted, resuspended in 10 mL of SC-His media containing 2% galactose, and used to inoculate a total volume of 100 mL of this media. Cultures were grown for 17 hours at 30°C and 250 rpm. Cells were then pelleted, rinsed in 0.85% NaCl, and repelleted. Cell pellets (190 mg) were suspended in 380 μL of 100 mM potassium phosphate buffer, pH 7.5, and an equal volume (pellet plus buffer) of pre-rinsed glass beads (Sigma, 150-212 microns) was added. This mixture was vortexed for 30 seconds and placed on ice for 1 minute, and the vortexing/cooling cycle was repeated 7 additional times. The cells were then centrifuged for 6 minutes at 5,000 *g* and the supernatant was removed to a fresh tube. The beads/pellet were washed twice with 300 μL of buffer and centrifuged, and the supernatants joined with the first supernatant.

An anaerobically-grown culture of *E. coli* strain DH10B was used as a positive control for D-LDH assays. A culture of this strain was grown to saturation overnight and diluted 1:20 the following morning in fresh LB media. The culture was grown anaerobically at 37°C for 7.5 hours. Cells were pelleted, washed with 0.85 % NaCl, and repelleted. Cells were disrupted using BugBuster™ Protein Extraction Reagent and Benzonase® (Novagen) as per manufacturer's instructions with a 20-minute incubation at room temperature. After centrifugation at 16,000g and 4°C, the supernatant was transferred to a new tube and used in the activity assay.

Total protein content of cell extracts was quantified using a microplate Bio-Rad Protein Assay (Bio-Rad, Hercules, CA). The D-LDH and D-LDH/PCT constructs in YPH500, the pESC-His negative control in YPH500, and the anaerobically-grown *E. coli* strain were tested for their ability to catalyze the conversion of pyruvate to lactate by assaying the concurrent oxidation of NADH to NAD. The assay mixture contained 100 mM potassium phosphate buffer, pH 7.5, 0.2 mM NADH, and 0.5-1.0 µg of cell extract. The reaction was started by the addition of sodium pyruvate to a final concentration of 5 mM, and the decrease in absorbance at 340 nm was measured over 10 minutes. When 0.5 µg of total cell extract protein of the negative control strain was added to the reaction mixture, a decrease in absorbance from -0.01 to –0.02 was seen over 200 seconds. A decrease in absorbance from –0.21 to –0.47 and –0.20 to –0.47 over 200 seconds was seen, respectively, for cell extract from the YPH500/pESC-His/D-LDH or YPH500/pESC-His/D-LDH/PCT strains. 0.5 µL (7.85 µg) of cell extract from the anaerobically-grown *E. coli* strain showed a decrease in absorbance very similar to that for 1 µg of cell extract of the YPH500/pESC-His/D-LDH/PCT strain. When 4 µg of cell extract was used, the YPH500/pESC-His/D-LDH/PCT strain showed a decrease in absorbance from –0.18 to –0.60 over 10 minutes, whereas the negative control strain showed no decrease in absorbance (-0.08 to –0.08).

D. Demonstration of 3-HP production in *S. cerevisiae*

The pESC-Trp/OS19/EI, pESC-Leu/EIIa/EIIB, and pESC-His/D-LDH/PCT constructs were transformed into a single strain of *S. cerevisiae* YPH500 using a Frozen-EZ Yeast Transformation II™ Kit (Zymo Research, Orange, CA). A negative control

strain was also developed by transformation of the pESC-Trp, pESC-Leu, and pESC-His vectors into a single YPH500 strain. Transformation reactions were plated on SC-Trp-Leu-His media. Individual yeast colonies were screened by colony PCR for the presence or absence of nucleic acid corresponding to each construct.

The strain carrying all six genes and the negative control strain were grown in 5 mL of SC-Trp-Leu-His media containing 2% glucose. These cultures were grown for 31 hours at 30°C, and 2 mL was used to inoculate 50 mL of the same media. The subcultures were grown for 19 hours at 30°C, and their OD600s were determined. For each strain, a volume of cells giving an OD x volume equal to 100 was pelleted, washed with SC-Trp-Leu-His media with no carbon source, and repelleted. The cells were suspended in 10 mL of SC-Trp-Leu-His media containing 2% galactose and 2% raffinose and used to inoculate a total volume of 250 mL of this media. The cultures were grown in bottles at 30°C with no shaking, and samples were taken at 0, 4.5, 20, 28.5, 45, and 70 hours. Samples were spun down to remove cells and the supernatant was filtered using 0.45 micron Acrodisc Syrige Filters (Pall Gelman Laboratory, Ann Arbor, MI).

100 microliters of the filtered broth was used to derive CoA esters of any lactate or 3-HP in the broth using 6 micrograms of purified propionyl-CoA transferase, 50 mM potassium phosphate buffer (pH 7.0), and 1 mM acetyl-CoA. The reaction was allowed to proceed at room temperature for 15 minutes and was stopped by adding 1 volume 10% trifluoroacetic acid. The reaction mixtures were purified using Sep Pak C18 columns as previously described and analyzed by LC/MS.

## Example 14 Constructing a Biosynthetic Pathway that Produces Organic Acids from β-alanine

One possible pathway to 3-HP from β-alanine involves the use of a polypeptide having CoA transferase activity (e.g., an enzyme from a class of enzymes that transfers a CoA group from one metabolite to the other). As shown in Figure 54, β-alanine can be converted to β-alanyl-CoA using a polypeptide having CoA transferase activity and CoA donors such as acetyl-CoA or propionyl-CoA. Alternatively, β-alanyl-CoA can be generated by the action of a polypeptide having CoA synthetase activity. The β-alanyl-CoA can be deaminated to form acrylyl-CoA by a polypeptide having β-alanyl-CoA

ammonia lyase activity. The hydration of acrylyl-CoA at the β position to yield 3-HP-CoA can be carried out by a polypeptide having 3-HP-CoA dehydratase activity. The 3-HP-CoA can act as a CoA donor for β-alanine, a reaction that can be catalyzed a polypeptide having CoA transferase activity, thus yielding 3-HP as a product. Alternatively, 3-HP-CoA can be hydrolyzed to yield 3-HP by a polypeptide having specific CoA hydrolase activity.

Methods for isolating, sequencing, expressing, and testing the activity of a polypeptide having CoA transferase activity are described herein.

A. Isolation of a polypeptide having β-alanyl-CoA Ammonia Lyase Activity

Polypeptides having β-alanyl-CoA ammonia lyase activity can catalyze the conversion of β-alanyl-CoA into acryly-CoA. The activity of such polypeptides has been described by Vagelos *et al.* (*J. Biol. Chem.*, 234:490-497 (1959)) in *Clostridum propionicum*. This polypeptide can be used as part of the acrylate pathway in *Clostridum propionicum* to produce propionic acid.

*C. propionicum* was grown at 37°C in an anoxic medium containing 0.2% yeast extract, 0.2% trypticase peptone, 0.05% cysteine, 0.5% b-alanine, 0.4% VRB-salts, 5 mM potassium phosphate, pH 7.0. The cells were harvested after 12 hours and washed twice with 50 mM potassium phosphate (Kpi), pH 7.0. About 2 g of wet packed cells were re-suspended in 40 mL of Kpi, pH 7.0, 1mM $MgCl_2$, 1 mM EDTA, and 1 mM DTT (Buffer A), and homogenized by sonication at about 85-100 W power using a 3mm tip (Branson sonifier 250). Cell debris was removed by centrifugation at 100,000*g* for 45 minutes in a Centricon T-1080 Ultra centrifuge, and the cell free extract ( ~ 110 U/mg activity) was subjected to anion exchange chromatography on Source-15Q-material. The Source-15Q column was loaded with 32 mL of cell free extract. The column was developed by a linear gradient of 0 M to 0.5 M NaCl within 10 column volumes. The polypeptide eluted between 70-110 mM NaCl.

The solution was adjusted to a final concentration of 1 M $(NH_4)_2SO_4$ and applied onto a Resource-Phe column equilibrated with 1 M $(NH_4)_2SO_4$ in buffer A. The polypeptide did not bind to this column.

The final preparation was obtained after concentration in an Amicon chamber (filter cut-off 30 kDa). The functional polypeptide is composed of four polypeptide sub-units, each having a molecular mass of 16 kDa. The polypeptide had a final specific activity of 1033 U/mg in the standard assay (see below).

The polypeptide sample after every purification step was separated on a 15% SDS-PAGE gel. The gel was stained with 0.1% Coomassie R 250, and the destaining was achieved by using 7.1% acetic acid/5% ethanol solution.

The polypeptide was desalted by RP-HPLC and subjected to N-terminal sequencing by gas phase Edman degradation. The results of this analysis yielded a 35 amino acid N-terminal sequence of the polypeptide. The sequence was as follows: MV-GKKVVHHLMMSAKDAHYTGNLVNGARIVNQWGD (SEQ ID NO:177).

B. Amplification of a Gene Fragment

The 35 amino acid sequence of the polypeptide having β-alanine-CoA ammonia lyase activity was used to design primers with which to amplify the corresponding DNA from genome of *C. propionicium*. Genomic DNA from *C. propionicum* was isolated using the Gentra Genomic DNA isolation Kit (Gentra Systems, Minneapolis) following the genomic DNA protocol for gram-positive bacteria. A codon usage table for *Clostridium propionicum* was used to back translate the seven amino acids on either end of the amino acid sequence to obtain 20-nucleotide degenerate primers:

ACLF: 5'-ATGGTWGGYAARAARGTWGT -3' (SEQ ID NO:178)

ACLR: 5'- TCRCCCCAYTGRTTWACRAT -3'(SEQ ID NO:179)

The primers were used in a 50 μL PCR reaction containing 1X Taq PCR buffer, 0.6 μM each primer, 0.2 mM each dNTP, 2 units of Taq DNA polymerase (Roche Molecular Biochemicals, Indianapolis, IN), 2.5% (v/v) DMSO, and 100 ng of genomic DNA. PCR was conducted using a touchdown PCR program with 4 cycles at an annealing temperature of 58°C, 4 cycles at 56°C, 4 cycles at 54°C, and 24 cycles at 52°C. Each cycle used an initial 30 second denaturing step at 94°C and a 1.25 minute extension at 72°C, and the program had an initial denaturation step at 94°C for 2 minutes and final extension at 72°C for 5 minutes. The amounts of PCR primer used in the reaction were increased three-fold above typical PCR amounts due to the amount of degeneracy in the

3' end of the primer. In addition, separate PCR reactions containing each individual primer were made to identify PCR product resulting from single degenerate primers. Twenty μL of each PCR product was separated on a 2.0% TAE (Tris-acetate-EDTA)-agarose gel.

A band of about 100 bp was produced by the reaction containing both the forward and reverse primers, but was not present in the individual forward and reverse primer control reactions. This fragment was excised and purified using a QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). Four microliters of the purified band was ligated into pCRII-TOPO vector and transformed by a heat-shock method into TOP10 *E. coli* cells using a TOPO cloning procedure (Invitrogen, Carlsbad, CA). Transformations were plated on LB media containing 50 μg/mL of kanamycin and 50 μg/mL of 5-Bromo-4-Chloro-3-Indolyl-B-D-Galactopyranoside (X-gal). Individual, white colonies were resuspended in 25 μL of 10 mM Tris and heated for 10 minutes at 95°C to break open the bacterial cells. Two microliters of the heated cells were used in a 25 μL PCR reaction using M13R and M13F universal primers homologous to the pCRII-TOPO vector. The PCR mix contained the following: 1X Taq PCR buffer, 0.2 μM each primer, 0.2 mM each dNTP, and 1 unit of Taq DNA polymerase per reaction. The PCR reaction was performed in a MJ Research PTC100 under the following conditions: an initial denaturation at 94°C for 2 minutes; 30 cycles of 94°C for 30 seconds, 52°C for 1 minute, and 72°C for 1.25 minutes; and a final extension for 7 minutes at 72°C. Plasmid DNA was obtained (QIAprep Spin Miniprep Kit, Qiagen) from cultures of colonies showing the desired insert and was used for DNA sequencing with M13R universal primer. The following nucleic acid sequence was internal to the degenerate primers and corresponds to a portion of the 35 amino acid residue sequence: 5'-ACATCATTTAATGATGA-GCGCAAAAGATGCTCACTATACTGGAAACTTAGTAAACGGCGCTAGA-3' (SEQ ID NO:180).

C. Genome Walking to Obtain the Complete Coding Sequence

Primers for conducting genome walking in both upstream and downstream directions were designed using the portion of the nucleic acid sequence that was internal to the degenerate primers. The primer sequences were as follows:

ACLGSP1F: 5'-GTACATCATTTAATGATGAGCGCAAAAGATG-3' (SEQ ID NO:181)

ACLGSP2F: 5'-GATGCTCACTATACTGGAAACTTAGTAAAC-3' (SEQ ID NO:182)

ACLGSP1R: 5'-ATTCTAGCGCCGTTTACTAAGTTTCCAG-3' (SEQ ID NO:183)

ACLGSP2R: 5'-CCAGTATAGTGAGCATCTTTTGCGCTCATC-3' (SEQ ID NO:184)

GSP1F and GSP2F are primers facing downstream, GSP1R and GSP2R are primers facing upstream, and GSP2F and GSP2R are primers nested inside GSP1F and GSP1R, respectively. Genome walking libraries were constructed according to the manual for CLONTECH's Universal Genome Walking Kit (CLONTECH Laboratories, Palo Alto, CA), with the exception that the restriction enzymes *Ssp* I and *Hinc* II were used in addition to *Dra* I, *EcoR* V, and *Pvu* II. PCR was conducted in a Perkin Elmer 9700 Thermocycler using the following reaction mix: 1X XL Buffer II, 0.2 mM each dNTP, 1.25 mM $Mg(OAc)_2$ , 0.2 μM each primer, 2 units of rTth DNA polymerase XL (Applied Biosystems, Foster City, CA), and 1 μL of library per 50 μL reaction. First round PCR used an initial denaturation at 94°C for 5 seconds; 7 cycles consisting of 2 sec at 94°C and 3 min at 70°C; 32 cycles consisting of 2 sec at 94°C and 3 min at 64°C; and a final extension at 64°C for 4 min. Second round PCR used an initial denaturation at 94°C for 15 seconds; 5 cycles consisting of 5 sec at 94°C and 3 min at 70°C; 26 cycles consisting of 5 sec at 94°C and 3 min at 64°C; and a final extension at 66°C for 7 min. Twenty μL of each first and second round product was run on a 1.0% TAE-agarose gel. In the second round PCR for the forward reactions, a 1.4 Kb band was obtained for *Dra* I, a 1.5 Kb band for *Hinc* II, a 4.0 Kb band for *Pvu* II, and 2.0 and 2.6 Kb bands were obtained for *Ssp* I. In the second round PCR for the reverse reactions, a 1.5 Kb band was obtained for *Dra* I, a 0.8 Kb band for *EcoR* V, a 2.0 Kb band for *Hinc* II, a 2.9 Kb band for *Pvu* II, and a 1.5 Kb band was obtained for *Ssp* I. Several of these fragments were gel purified, cloned, and sequenced.

The coding sequence of the polypeptide having β-alanyl-CoA ammonia lyase activity is set forth in SEQ ID NO:162. This coding sequence encodes the amino acid sequence set forth in SEQ ID NO:160. The coding sequence was cloned and expressed in

bacterial cells. A polypeptide with the expected size was isolated and tested for enzymatic activity.

The isolation of a nucleic acid molecule encoding a polypeptide having 3-HP-CoA dehydratase activity (e.g., the seventh enzymatic activity in Figure 54, which can be accomplished with a polypeptide having the amino acid sequence set forth in SEQ ID NO:41) is described herein. This polypeptide in combination with a polypeptide having CoA transferase activity (e.g., a polypeptide having the amino acid sequence set forth in SEQ ID NO:2) and a polypeptide having β-alanyl-CoA ammonia lyase activity (e.g., a polypeptide having the amino acid sequence set forth in SEQ ID NO: 160) provides one method of making 3-HP from β-alanine.

## Example 15 Constructing a Biosynthetic Pathway that Produces Organic Acids from β-alanine

In another pathway, β-alanine generated from aspartate can be deaminated by a polypeptide having 4, 4-aminobutyrate aminotransferase activity (Figure 55). This reaction also can regenerate glutamate that is consumed in the formation of aspartate. The deamination of β-alanine can yield malonate semialdehyde, which can be further reduced to 3-HP by a polypeptide having 3-hydroxypropionate dehydrogenase activity or a polypeptide having 3-hydroxyisobutyrate dehydrogenase activity. Such polypeptides can be obtained as follows.

A. Cloning gabT (4-aminobutyrate aminotransferase) from *C. acetobutycilicum*

The following PCR primers were designed based on a published sequence for a gabT gene from *Clostridium acetobutycilicum* (GenBank# AE007654):

Cac aba nco sen: 5'-GAGCCATGGAAGAAATAAATGCTAAAG- 3' (SEQ ID NO:185)
Cac aba bam anti: 5'-AGAGGATGGCTTTTTAAATCGCTATTC- 3' (SEQ ID NO:186)

The primers introduced a *Nco*I site at the 5' end and a *BamH* I site at the 3' end. A PCR reaction was set up using chromosomal DNA from *C. acetobutylicum* as the template.

| | | |
|---|---|---|
| H2O | 80.75 µL | PCR Program |
| Taq Plus Long 10x Buffer | 10 µL | 94° C 5 minutes |
| dNTP mix (10 mM) | 3 µL | 25 cycles of: |
| Cac aba nco sen (20 mM) | 2 µL | 94° C 30 seconds |
| Cac aba bam anti (20 mM) | 2 µL | 50° C 30 seconds |
| C. acetobutylicum DNA (~100 ng) | 1 µL | 72° C 80 seconds + 2 |
| Taq Plus Long (5 U/mL) | 1 µL | seconds/cycle |
| Pfu (2.5 U/mL) | 0.25 µL | 1 cycle of : |
| | | 68° C 7 minutes |
| | | 4°C until use |

Upon agarose gel analysis a single band was observed of ~1.3 Kb in size. This fragment was purified using QIAquick PCR purification kit (Qiagen, Valencia, CA) and cloned into pCRII TOPO using the TOPO Zero Blunt PCR cloning kit (Invitrogen, Carlsbad, CA). 1 µL of the pCRII TOPO ligation mix was used to transform chemically competent TOP10 *E. coli* cells. The cells were for 1 hour in SOC media, and the transformants were selected on LB/kanamycin (50 µg/mL) plates. Single colonies of the transformant grown overnight in LB/kanamycin media, and the plasmid DNA was extracted using a Mini prep kit (Qiagen, Valencia, CA). The super-coiled plasmid DNA was separated on a 1% agarose gel digested, and the colonies with insert were selected. The insert was sequenced to confirm the sequence and its quality.

The plasmid having the correct insert was digested with restriction enzyme *Nco* I and *BamH* I. The digested insert was gel isolated and ligated to pET28b expression vector that was also restricted with *Nco* I and *BamH* I enzymes. 1 µl of ligation mix was used to transform chemically competent TOP10 *E. coli* cells. The cells were recovered for 1 hour in SOC media, and the transformants were selected on LB/kanamycin (50 µg/mL) plates. The super-coiled plasmid DNA was separated on a 1% agarose gel digested, and the colonies with insert were selected. The plasmid with the insert was isolated using a Mini prep kit (Qiagen, Valencia, CA), and 1 µL of this plasmid DNA was used to transform electrocompetent BL21(DE3) (Novagen, Madison, WI). These cells were used to check the expression of a polypeptide having 4-aminobutyrate aminotransferase activity.

B. Cloning mmsB (3-hydroxyisobutyrate dehydrogenase) from *P. aeruginosa*

The following PCR primers was designed based on a published sequence for a mmsB gene from *Pseudomona aeruginosa* (GenBank# M84911):

Ppu hid nde sen: 5'-ATACATATGACCGACCGACATCGCATT-3' (SEQ ID NO:186)

Ppu hid sal anti: 5'-ATAGTCGACGGGTCAGTCCTTGCCGCG-3' (SEQ ID NO:187)

The primers introduced a *Nde* I site at the 5' end and a *BamH* I site at the 3' end.

| | | |
|---|---|---|
| $H_2O$ | 80.75 µL | PCR Program |
| Taq Plus Long 10x Buffer | 10 µL | 94° C 5 minutes |
| dNTP mix (10 mM) | 3 µL | 25 cycles of:<br>94° C 30 seconds<br>55°C 30 seconds<br>72°C 90 seconds + 2 seconds/cycle |
| Ppu hid nde sen (20 µM) | 2 µL | 68°C 7 minutes |
| Ppu hid sal anti (20 µM) | 2 µL | 4° C until use |
| C. acetobutylicum DNA (~100 ng) | 1 µl | |
| Taq Plus Long (Stratagene, La Jolla, CA) | 1 µL | |
| Pfu (Stratagene, La Jolla, CA) | 0.25 µL | |

A PCR reaction was set up using chromosomal DNA from *P. aeruginosa* as the template. Chromosomal DNA was obtained from ATCC (Manassas, VA) *P. aeruginosa* 17933D.

Upon agarose gel analysis, a single band was observed of ~1.6 Kb in size. This fragment was purified using QIAquick PCR purification kit (Qiagen, Valencia, CA) and cloned into pCRII TOPO using the TOPO Zero Blunt PCR cloning kit (Invitrogen, Carlsbad, CA). 1 µL of the pCRII TOPO ligation mix was used to transform chemically competent TOP10 *E. coli* cells. The cells were recovered for 1 hour in SOC media, and the transformants were selected on LB/kanamycin (50 µg/mL) plates. Single colonies of the transformant grown overnight in LB/kanamycin media, and the plasmid DNA was extracted using a Mini prep kit (Qiagen, Valencia, CA). The super-coiled plasmid DNA

was separated on a 1% agarose gel and digested, and the colonies with insert were selected. The insert was sequenced to confirm the sequence and its quality.

The plasmid having the correct insert was digested with restriction enzyme *Nde* I and *BamHI*. The digested insert was gel isolated and ligated to pET30a expression vector that was also restricted with *Nde* I and *BamH* I enzymes. 1 μL of ligation mix was used to transform chemically competent TOP10 *E. coli* cells. The cells were recovered for 1 hour in SOC media, and the transformants were selected on LB/kanamycin (50 μg/mL) plates. The super-coiled plasmid DNA was separated on a 1% agarose gel and digested, and the colonies with insert were selected. The plasmid with the insert was isolated using a Mini prep kit (Qiagen, Valencia, CA), and 1 μl of this plasmid DNA was used to transform electrocompetent BL21(DE3) (Novagen, Madison, WI). These cells were used to check the expression of a polypeptide having 3-hydroxyisobutyrate dehydrogenase activity.

## OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

**WHAT IS CLAIMED IS:**

1. A cell comprising lactyl-CoA dehydratase activity and 3-hydroxypropionyl-CoA dehydratase activity.

2. The cell of claim 1, wherein said cell comprises an activity selected from the group consisting of E1 activator activity, E2 α activity, and E2 β activity.

3. The cell of claim 1, wherein said cell comprises 3-hydroxypropionyl-CoA dehydratase activity.

4. The cell of claim 1, wherein said cell comprises CoA transferase activity.

5. The cell of claim 1, wherein said cell comprises an exogenous nucleic acid comprising:

(a) a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163; or

(b) a nucleic acid sequence that shares at least 65 percent sequence identity with a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163.

6. The cell of claim 1, wherein said cell comprises 3-hydroxypropionyl-CoA hydrolase activity or 3-hydroxyisobutryl-CoA hydrolase activity.

7. The cell of claim 1, wherein said cell comprises lipase activity.

8. The cell of claim 1, wherein said cell produces 3-HP.

9. The cell of claim 1, wherein said cell produces an ester of 3-HP.

10. The cell of claim 9, wherein said ester is selected from the group consisting of methyl 3-hydroxypropionate, ethyl 3-hydroxypropionate, propyl 3-hydroxypropionate, butyl 3-hydroxypropionate, and 2-ethylhexyl 3-hydroxypropionate.

11. The cell of claim 1, wherein said cell comprises CoA synthetase activity.

12. The cell of claim 1, wherein said cell comprises poly hydroxyacid synthase activity.

13. The cell of claim 1, wherein said cell produces polymerized 3-HP.

14. The cell of claim 1, wherein said cell is prokaryotic.

15. The cell of claim 1, wherein said cell is selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

16. A cell comprising CoA synthetase activity, lactyl-CoA dehydratase activity, and poly hydroxyacid synthase activity.

17. The cell of claim 16, wherein said cell comprises an activity selected from the group consisting of E1 activator activity, E2 α activity, and E2 β activity.

18. The cell of claim 16, wherein the cell produces polymerized acrylate.

19. The cell of claim 16, wherein said cell is prokaryotic.

20. The cell of claim 16, wherein said cell is selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

21. A cell comprising CoA transferase activity, lactyl-CoA dehydratase activity, and lipase activity.

22. The cell of claim 21, wherein said cell comprises an activity selected from the group consisting of E1 activator activity, E2 α activity, and E2 β activity.

23. The cell of claim 21, wherein said cell produces an ester of acrylate.

24. The cell of claim 23, wherein said ester is selected from the group consisting of methyl acrylate, ethyl acrylate, propyl acrylate, and butyl acrylate.

25. The cell of claim 21, wherein said cell is prokaryotic.

26. The cell of claim 21, wherein said cell is selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

27. An polypeptide comprising an amino acid sequence selected from the group consisting of:

(a) a sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161;

(b) a sequence having at least 10 contiguous amino acid residues of a sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161;

(c) a sequence that has at least 65 percent sequence identity with a sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161;

(d) a sequence that has at least 65 percent sequence identity with at least 10 contiguous amino acid residues of a sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161; and

(e) a sequence set forth in SEQ ID NO:2, 10, 18, 26, 35, 37, 39, 41, 141, 160, or 161 that contains at least one conservative substitution.

28. A nucleic acid molecule comprising a nucleic acid sequence that encodes the polypeptide of claim 27.

29. A transformed cell comprising at least one exogenous nucleic acid molecule, wherein said molecule comprises a nucleic acid sequence that encodes the polypeptide of claim 27.

30. The cell of claim 29, wherein the cell produces 3-HP.

31. The cell of claim 29, wherein said exogenous nucleic acid molecule encodes an E2 α polypeptide of an enzyme having lactyl-CoA dehydratase activity.

32. The cell of claim 29, wherein said exogenous nucleic acid molecule encodes an E2 β polypeptide of an enzyme having said lactyl-CoA dehydratase activity.

33. The cell of claim 29, wherein said exogenous nucleic acid molecule encodes a polypeptide having 3-hydroxypropionyl-CoA dehydratase activity or CoA transferase activity.

34. The cell of claim 29, wherein said exogenous nucleic acid molecule encodes a polypeptide having 3-hydroxypropionyl-CoA hydrolase activity or 3-hydroxyisobutryl-CoA hydrolase activity.

35. The cell of claim 29, wherein the cell comprises lipase activity.

36. The cell of claim 29, wherein the cell produces an ester of 3-HP.

37. The cell of claim 36, wherein said ester is selected from the group consisting of methyl 3-hydroxypropionate, ethyl 3-hydroxypropionate, propyl 3-hydroxypropionate, butyl 3-hydroxypropionate, and 2-ethylhexyl 3-hydroxypropionate.

38. The cell of claim 29, wherein said cell comprises CoA synthetase activity.

39. The cell of claim 29, wherein said cell produces polymerized 3-HP.

40. The cell of claim 29, wherein said cell is prokaryotic.

41. The cell of claim 29, wherein said cell is selected from the group consisting of *Lactobacillus, Lactococcus, Bacillus*, and *Escherichia* cells.

42. The cell of claim 29, wherein the cell is a yeast cell.

43. A specific binding agent that specifically binds to the polypeptide of claim 27.

44. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of:

(a) a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163;

(b) a sequence having at least 10 contiguous nucleotides of a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163;

(c) a sequence that has at least 65 percent sequence identity with a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163;

(d) a sequence that has at least 65 percent sequence identity with at least 10 contiguous nucleotides of a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163; and

(e) a sequence that hybridize under moderately stringent conditions a sequence set forth in SEQ ID NO:1, 9, 17, 25, 33, 34, 36, 38, 40, 42, 129, 140, 142, 162, or 163.

45. A production cell comprising an isolated nucleic acid molecule of claim 44 that is exogenous to said production cell.

46. The cell of claim 45, wherein said isolated nucleic acid molecule encodes a polypeptide having an enzymatic activity selected from the group consisting of CoA transferase activity, lactyl-CoA dehydratase activity, CoA synthase activity, CoA

dehydratase activity, dehydrogenase activity, malonyl-CoA reductase activity, and 3-hydroxypropionyl-CoA dehydratase activity.

47. A method of producing a polypeptide, comprising culturing the cell of claim 45 under conditions that allow said cell to produce said polypeptide, wherein said polypeptide is produced.

48. A method for making 3-HP, said method comprising culturing at least one cell comprising at least one exogenous nucleic acid molecule that encodes at least one polypeptide that is capable of producing said 3-HP from PEP under conditions such that said 3-HP is produced.

49. The method of claim 48, wherein said cell is selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

50. The method of claim 48, wherein 3-HP is made by a biosynthetic route that utilizes a β-alanine intermediate.

51. The method of claim 48, wherein 3-HP is made by a biosynthetic route that utilizes a malonyl-CoA intermediate.

52. The method of claim 48, wherein 3-HP is made by a biosynthetic route that utilizes a lactate intermediate.

53. A method for making 3-HP, said method comprising culturing at least one cell comprising at least one exogenous nucleic acid molecule that encodes at least one polypeptide that is capable of producing said 3-HP from lactate under conditions such that said 3-HP is produced.

54. The method of claim 53, wherein said cells are selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

55. A method for making 3-HP, said method comprising culturing at least one cell under conditions wherein said cell produces said 3-HP, said cell comprising lactyl-CoA dehydratase activity and 3-hydroxypropionyl-CoA dehydratase activity.

56. The method of claim 55, wherein said cell is selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

57. The method of claim 55, wherein said cell comprises CoA transferase activity.

58. The method of claim 55, wherein said cell comprises 3-hydroxypropionyl-CoA hydrolase activity or 3-hydroxyisobutryl-CoA hydrolase activity.

59. A method for making 3-HP, said method comprising:

a) contacting lactate with a first polypeptide having CoA transferase activity to form lactyl-CoA,

b) contacting said lactyl-CoA with a second polypeptide having lactyl-CoA dehydratase activity to form acrylyl-CoA,

c) contacting said acrylyl-CoA with a third polypeptide having 3-hydroxypropionyl-CoA dehydratase activity to form 3-HP-CoA, and

d) contacting said 3-HP-CoA with said first polypeptide to form said 3-HP or with a fourth polypeptide having 3-hydroxypropionyl-CoA hydrolase activity or 3-hydroxyisobutryl-CoA hydrolase activity to form said 3-HP.

60. A method for making polymerized 3-HP, said method comprising culturing a cell under conditions wherein said cell produces said polymerized 3-HP, said cell comprising lactyl-CoA dehydratase activity and 3-hydroxypropionyl-CoA dehydratase activity.

61. The method of claim 60, wherein said cell is selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

62. The method of claim 60, wherein said cell comprises CoA synthetase activity.

63. The method of claim 60, wherein said cell comprises poly hydroxyacid synthase activity.

64. A method for making polymerized 3-HP, said method comprising:

a) contacting lactate with a first polypeptide having CoA synthetase activity to form lactyl-CoA,

b) contacting said lactyl-CoA with a second polypeptide having lactyl-CoA dehydratase activity to form acrylyl-CoA,

c) contacting said acrylyl-CoA with a third polypeptide having 3-hydroxypropionyl-CoA dehydratase activity to form 3-hydroxypropionic acid-CoA, and

d) contacting said 3-hydroxypropionic acid-CoA with a fourth polypeptide having poly hydroxyacid synthase activity to form said polymerized 3-HP.

65. A method for making an ester of 3-HP, said method comprising culturing a cell under conditions wherein said cell produces said ester, said cell comprising lactyl-CoA dehydratase activity and 3-hydroxypropionyl-CoA dehydratase activity.

66. The method of claim 65, wherein said cell is selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

67. The method of claim 65, wherein said cell comprises CoA transferase activity.

68. The method of claim 65, wherein said cell comprises 3-hydroxypropionyl-CoA hydrolase activity or 3-hydroxyisobutryl-CoA hydrolase activity.

69. A method for making an ester of 3-HP, said method comprising:

a) contacting lactate with a first polypeptide having CoA transferase activity to form lactyl-CoA,

b) contacting said lactyl-CoA with a second polypeptide having lactyl-CoA

dehydratase activity to form acrylyl-CoA,

c) contacting said acrylyl-CoA with a third polypeptide having 3-hydroxypropionyl-CoA dehydratase activity to form 3-hydroxypropionic acid-CoA,

d) contacting said 3-hydroxypropionic acid-CoA with said first polypeptide to form 3-HP or a fourth polypeptide having 3-hydroxypropionyl-CoA hydrolase activity or 3-hydroxyisobutryl-CoA hydrolase activity to form 3-HP, and

e) contacting said 3-HP with a fifth polypeptide having lipase activity to form said ester.

70. A method for making polymerized acrylate, said method comprising culturing a cell under conditions wherein said cell produces said polymerized acrylate, said cell comprising CoA synthetase activity and lactyl-CoA dehydratase activity.

71. The method of claim 70, wherein said cell is selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

72. The method of claim 70, wherein said cell comprises poly hydroxyacid synthase activity.

73. A method for making polymerized acrylate, said method comprising:

a) contacting lactate with a first polypeptide having CoA synthetase activity to form lactyl-CoA,

b) contacting said lactyl-CoA with a second polypeptide having lactyl-CoA dehydratase activity to form acrylyl-CoA, and

c) contacting said acrylyl-CoA with a third polypeptide having poly hydroxyacid synthase activity to form said polymerized acrylate.

74. A method for making an ester of acrylate, said method comprising culturing a cell under conditions wherein said cell produces said ester, said cell comprising CoA transferase activity and lactyl-CoA dehydratase activity.

75. The method of claim 74, wherein said cell is selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

76. The method of claim 74, wherein said cell comprises lipase activity.

77. A method for making an ester of acrylate, said method comprising:

a) contacting lactate with a first polypeptide having CoA transferase activity to form lactyl-CoA,

b) contacting said lactyl-CoA with a second polypeptide having lactyl-CoA dehydratase activity to form acrylyl-CoA,

c) contacting said acrylyl-CoA with said first polypeptide to form acrylate, and

d) contacting said acrylate with a third polypeptide having lipase activity to form said ester.

78. A method for making 3-HP, said method comprising culturing a cell under conditions wherein said cell produces said 3-HP, said cell comprising at least one exogenous nucleic acid that encodes at least one polypeptide such that said 3-HP is produced from acetyl-CoA and under conditions such that said 3-HP is produced.

79. The method of claim 78, wherein said cell is selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

80. A method for making 3-HP, said method comprising culturing a cell under conditions wherein said cell produces said 3-HP, said cell comprising at least one exogenous nucleic acid that encodes at least one polypeptide such that said 3-HP is produced from malonyl-CoA and under conditions such that said 3-HP is produced.

81. The method of claim 80, wherein said cell is selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

82. A method for making 3-HP, said method comprising culturing a cell under

conditions wherein said cell produces said 3-HP, said cell comprising at least one exogenous nucleic acid that encodes at least one polypeptide such that said 3-HP is produced from β-alanine and under conditions such that said 3-HP is produced.

83. The method of claim 82, wherein said cell is selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

84. A method for making 3-HP, said method comprising culturing cells comprising an exogenous nucleic acid that encodes polypeptides that are capable of producing 3-HP from acetyl-CoA under conditions such that said 3-HP is produced.

85. The method of claim 84, wherein said cells are selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

86. A method for making 3-HP, said method comprising culturing cells comprising at least one exogenous nucleic acid that encodes polypeptides that are capable of producing said 3-HP from malonyl-CoA, and under conditions such that said 3-HP is produced.

87. The method of claim 86, wherein said cells are selected from the group consisting of yeast, *Lactobacillus*, *Lactococcus*, *Bacillus*, and *Escherichia* cells.

88. A method for making 3-HP, said method comprising:

a) contacting acetyl-CoA with a first polypeptide having acetyl-CoA carboxylase activity to form malonyl-CoA, and

b) contacting said malonyl-CoA with a second polypeptide having malonyl-CoA reductase activity to form said 3-HP.

89. A method for making 3-HP, said method comprising contacting malonyl-CoA with a polypeptide having malonyl-CoA reductase activity to form said 3-HP.

90. A method for making 3-HP, said method comprising:

a) contacting β-alanine CoA with a first polypeptide having β-alanyl-CoA ammonia lyase activity to form acrylyl-CoA;

b) contacting said acrylyl-CoA with a second polypeptide having 3HP-CoA dehydratase activity to form said 3-HP-CoA; and

c) contacting 3-HP-CoA with a third polypeptide having glutamate dehydrogenase to make 3-HP.

**Figure 1**

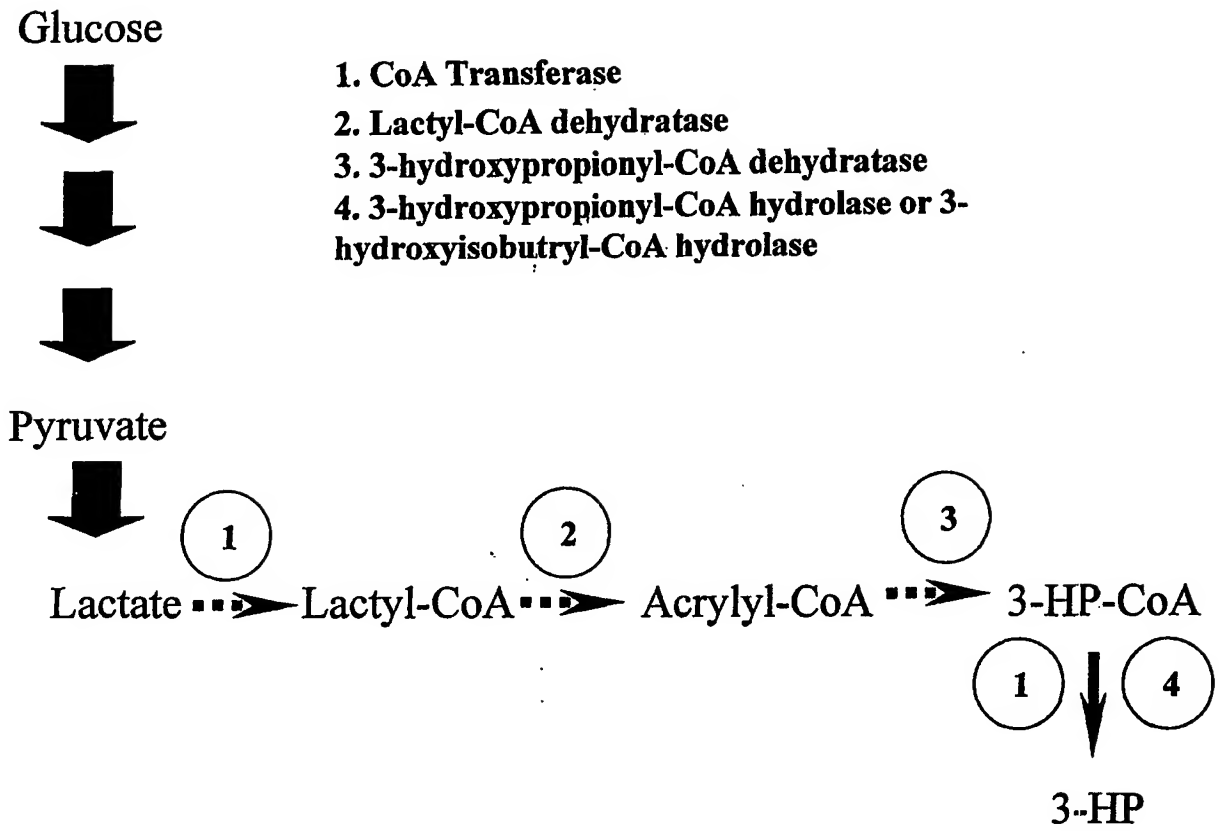

Glucose
1. CoA Transferase
2. Lactyl-CoA dehydratase
3. 3-hydroxypropionyl-CoA dehydratase
4. 3-hydroxypropionyl-CoA hydrolase or 3-hydroxyisobutryl-CoA hydrolase
Pyruvate
1
2
3
Lactate
Lactyl-CoA
Acrylyl-CoA
3-HP-CoA
1
4
3-HP

Figure 2

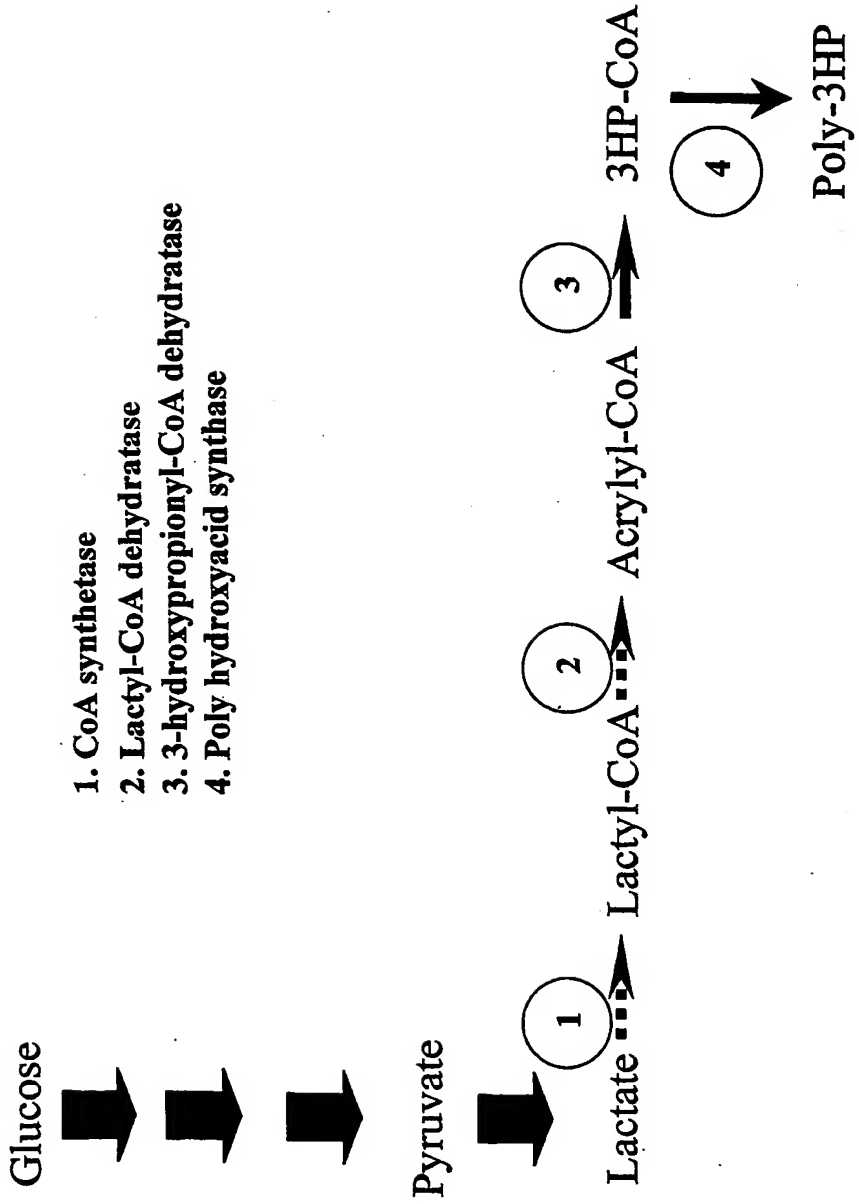

1. CoA synthetase
2. Lactyl-CoA dehydratase
3. 3-hydroxypropionyl-CoA dehydratase
4. Poly hydroxyacid synthase
Glucose
Pyruvate
Lactate
1
Lactyl-CoA
2
Acrylyl-CoA
3
3HP-CoA
4
Poly-3HP

**Figure 3**

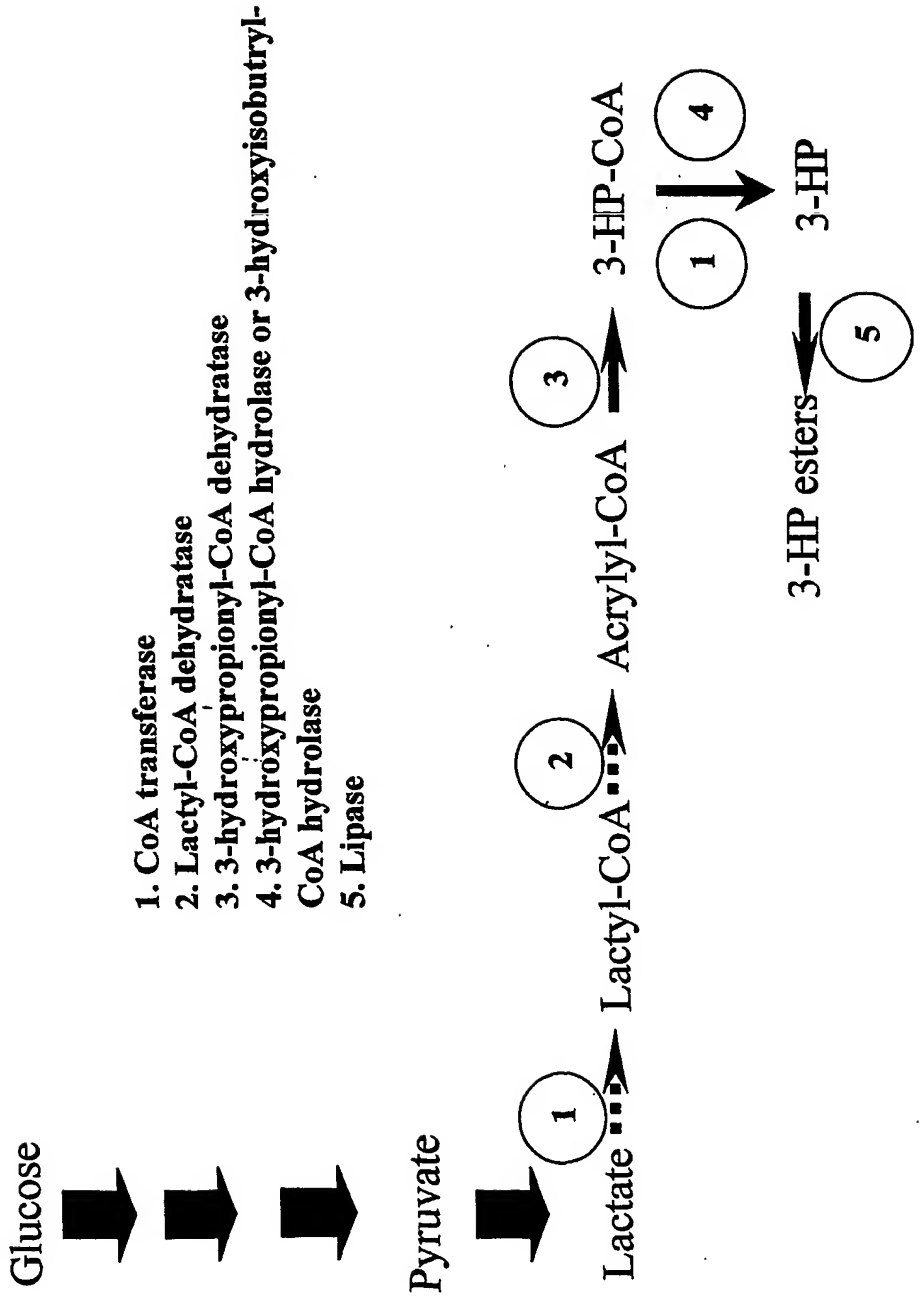

Glucose
Pyruvate
Lactate
1
Lactyl-CoA
2
Acrylyl-CoA
3
3-HP-CoA
1
4
3-HP
5
3-HP esters
1. CoA transferase
2. Lactyl-CoA dehydratase
3. 3-hydroxypropionyl-CoA dehydratase
4. 3-hydroxypropionyl-CoA hydrolase or 3-hydroxyisobutryl-CoA hydrolase
5. Lipase

Figure 4

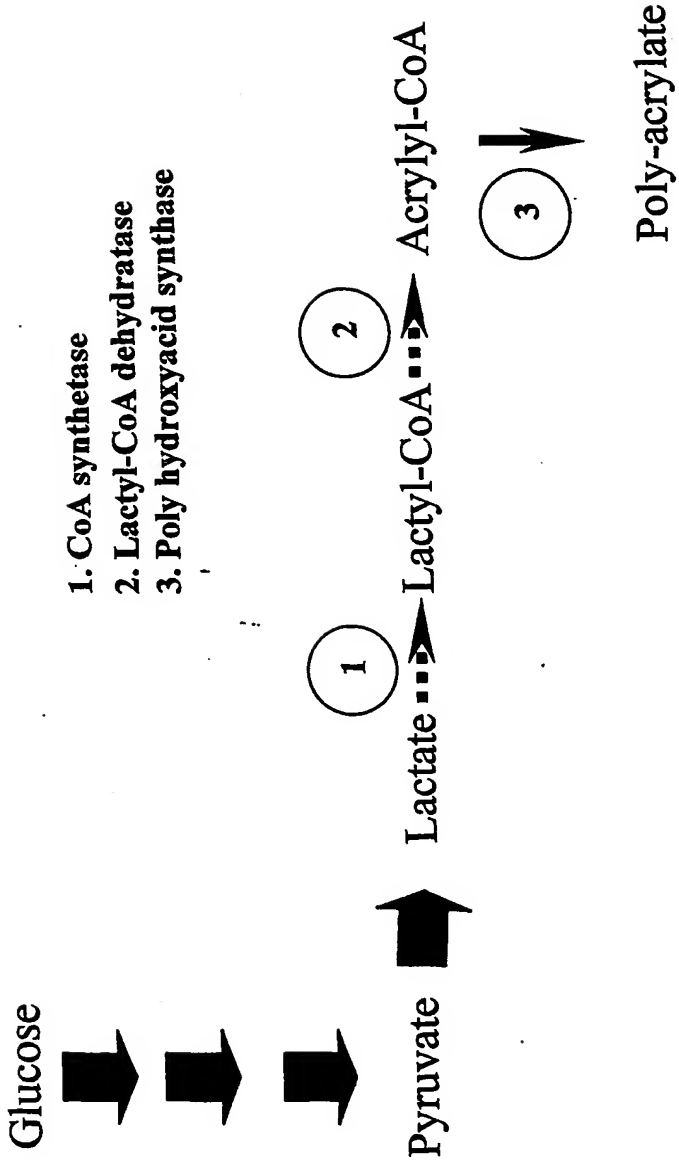

1. CoA synthetase
2. Lactyl-CoA dehydratase
3. Poly hydroxyacid synthase
Glucose
Pyruvate
Lactate
Lactyl-CoA
Acrylyl-CoA
Poly-acrylate
1
2
3

Figure 5

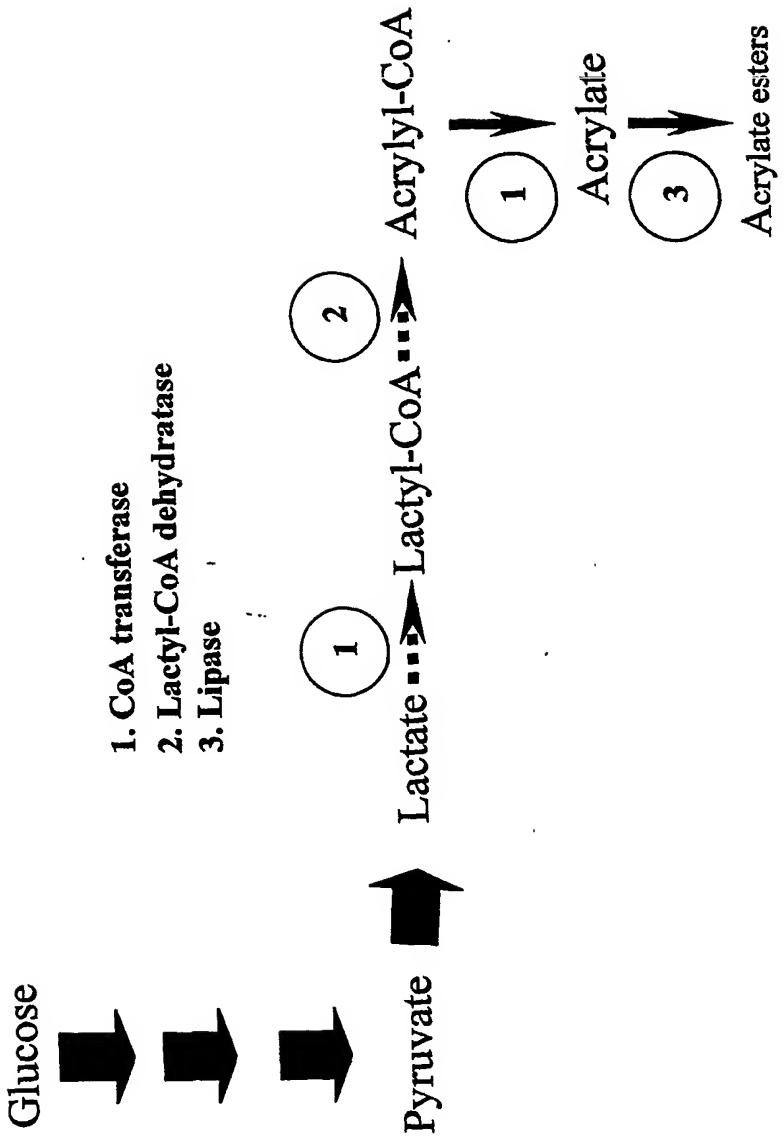

Glucose
Pyruvate
Lactate
Lactyl-CoA
Acrylyl-CoA
Acrylate
Acrylate esters
1
2
1
3
1. CoA transferase
2. Lactyl-CoA dehydratase
3. Lipase

**Figure 6**

ATGAGAAAAGTAGAAATCATTACAGCTGAACAAGCAGCTCAGCTCGTAAAAGACAACGAC
ACGATTACGTCTATCGGCTTTGTCAGCAGCGCCCATCCGGAAGCACTGACCAAAGCTTTG
GAAAAACGGTTCCTGGACACGAACACCCCGCAGAACTTGACCTACATCTATGCAGGCTCT
CAGGGCAAACGCGATGGCCGTGCCGCTGAACATCTGGCACACACAGGCCTTTTGAAACGC
GCCATCATCGGTCACTGGCAGACTGTACCGGCTATCGGTAAACTGGCTGTCGAAAACAAG
ATTGAAGCTTACAACTTCTCGCAGGGCACGTTGGTCCACTGGTTCCGCGCCTTGGCAGGT
CATAAGCTCGGCGTCTTCACCGACATCGGTCTGGAAACTTTCCTCGATCCCCGTCAGCTC
GGCGGCAAGCTCAATGACGTAACCAAAGAAGACCTCGTCAAACTGATCGAAGTCGATGGT
CATGAACAGCTTTTCTACCCGACCTTCCCGGTCAACGTAGCTTTCCTCCGCGGTACGTAT
GCTGATGAATCCGGCAATATCACCATGGACGAAGAAATCGGGCCTTTCGAAAGCACTTCC
GTAGCCCAGGCCGTTCACAACTGTGGCGGTAAAGTCGTCGTCCAGGTCAAAGACGTCGTC
GCTCACGGCAGCCTCGACCCGCGCATGGTCAAGATCCCTGGCATCTATGTCGACTACGTC
GTCGTAGCAGCTCCGGAAGACCATCAGCAGACGTATGACTGCGAATACGATCCGTCCCTC
AGCGGTGAACATCGTGCTCCTGAAGGCGCTACCGATGCAGCTCTCCCCATGAGCGCTAAG
AAAATCATCGGCCGCCGCGGCGCTTTGGAATTGACTGAAAACGCTGTCGTCAACCTCGGC
GTCGGTGCTCCGGAATACGTTGCTTCTGTTGCCGGTGAAGAAGGTATCGCCGATACCATT
ACCCTGACCGTCGAAGGTGGCGCCATCGGTGGCGTACCGCAGGGCGGTGCCCGCTTCGGT
TCGTCCCGCAATGCCGATGCCATCATCGACCACACCTATCAGTTCGACTTCTACGATGGC
GGCGGTCTGGACATCGCTTACCTCGGCCTGGCCCAGTGCGATGGCTCGGGCAACATCAAC
GTCAGCAAGTTCGGTACTAACGTTGCCGGCTGCGGCGGTTTCCCCAACATTTCCCAGCAG
ACACCGAATGTTTACTTCTGCGGCACCTTCACGGCTGGCGGCTTGAAAATCGCTGTCGAA
GACGGCAAAGTCAAGATCCTCCAGGAAGGCAAAGCCAAGAAGTTCATCAAAGCTGTCGAC
CAGATCACTTTCAACGGTTCCTATGCAGCCCGCAACGGCAAACACGTTCTCTACATCACA
GAACGCTGCGTATTTGAACTGACCAAAGAAGGCTTGAAACTCATCGAAGTCGCACCGGGC
ATCGATATTGAAAAAGATATCCTCGCTCACATGGACTTCAAGCCGATCATTGATAATCCG
AAACTCATGGATGCCCGCCTCTTCCAGGACGGTCCCATGGGACTGAAAAAATAA (SEQ
ID NO:1)

**Figure 7**

MRKVEIITAEQAAQLVKDNDTITSIGFVSSAHPEALTKALEKRFLDTNTPQNLTYIYAGS
QGKRDGRAAEHLAHTGLLKRAIIGHWQTVPAIGKLAVENKIEAYNFSQGTLVHWFRALAG
HKLGVFTDIGLETFLDPRQLGGKLNDVTKEDLVKLIEVDGHEQLFYPTFPVNVAFLRGTY
ADESGNITMDEEIGPFESTSVAQAVHNCGGKVVVQVKDVVAHGSLDPRMVKIPGIYVDYV
VVAAPEDHQQTYDCEYDPSLSGEHRAPEGATDAALPMSAKKIIGRRGALELTENAVVNLG
VGAPEYVASVAGEEGIADTITLTVEGGAIGGVPQGGARFGSSRNADAIIDHTYQFDFYDG
GGLDIAYLGLAQCDGSGNINVSKFGTNVAGCGGFPNISQQTPNVYFCGTFTAGGLKIAVE
DGKVKILQEGKAKKFIKAVDQITFNGSYAARNGKHVLYITERCVFELTKEGLKLIEVAPG
IDIEKDILAHMDFKPIIDNPKLMDARLFQDGPMGLKK (SEQ ID NO:2)

## Figure 8

```
SEQ ID NO:1       1 atgagaaaagtagaaatcattacagctgaacaagcagctc--agctcgta
SEQ ID NO:3       1 ---------gtgccggtcctgtcggcacaggaagcggtga--attatatt
SEQ ID NO:4       1 atgccgattctctcaaaaatatgggcggctccagcagctggaatcttgag
SEQ ID NO:5       1 -------------------------atgaa----------------tgca

SEQ ID NO:1      49 aaagacaacgacacgattacgtctatcggctttgtcagcagcgcccatcc
SEQ ID NO:3      40 cccgacgaagcaacactttgtgtgttaggcgctg---gcggcggtattct
SEQ ID NO:4      51 aaaaactccgagaaatgctcatcaaatgaggctaatctcaatga-catcc
SEQ ID NO:5      10 aaaga----------atta-----atcg----------------------

SEQ ID NO:1      99 ggaagcactgaccaaagctttggaaaaacggttcctg-------------
SEQ ID NO:3      87 ggaag------ccaccacgtt--aattactgctcttgctgataaatataa
SEQ ID NO:4     100 tcgatgaaagcaaaagtcttt-----aactctgc---------------
SEQ ID NO:5      23 ------------------------------------------------

SEQ ID NO:1     136 ---gacacgaacaccccgcagaacttgacctacatctatgcag-gctctc
SEQ ID NO:3     129 acagactcaaacaccacgt--aatttatcgattattagtccaa-cagggc
SEQ ID NO:4     129 -------cgaagaagccgtgaaggatattccagat-aatgcaaagctttt
SEQ ID NO:5      23 --------------ctcgccgaatt-------------------------

SEQ ID NO:1     182 agggcaaacgcgatggccgtgccgctgaacatctggcacacacaggcctt
SEQ ID NO:3     176 ttggcgatcgcgccgaccgtggtattagtcctctggcgcaagaaggtctg
SEQ ID NO:4     171 a----------gttggc--ggcttcggactatgcgg-aatcccagaaaat
SEQ ID NO:5      34 ---------gcgatgg-----------------------------------

SEQ ID NO:1     232 ttgaaacgcgccatcatcggtcactggcagactgtaccggc-tatcggta
SEQ ID NO:3     226 gtgaaatgggcattatgtggtcactgg-ggacaatcgccgcgtatttctg
SEQ ID NO:4     208 ctcatccaagctatca-caaaaactggtcaa-----------aaaggtc
SEQ ID NO:5      41 ----------------------aattacatgatgga---ga-tattgtta

SEQ ID NO:1     281 aactggctgtcgaaaacaagattgaagcttacaacttctcgcagggcacg
SEQ ID NO:3     275 aactcgcagaacaaaataaaattattgcttataactacccacaaggtgta
SEQ ID NO:4     245 ttacatgtgtatcaaacaatgcgggagttgataatt-------ggggac-
SEQ ID NO:5      65 atctcggt-----------attg--gtttac---------caacacagg

SEQ ID NO:1     331 ttggtccactggttccgcgccttggcaggtcataagctcggcgtcttcac
SEQ ID NO:3     325 cttacacaaaccttacgcgccgccgcagcccaccagcctggtattattag
SEQ ID NO:4     287 ttggcttgctccttc--aaactcgacaaatc--aagaaaatgatctcatc
SEQ ID NO:5      92 ttgt-----taattatttacctgataatgtcaata-----------ttac

SEQ ID NO:1     381 cgacatcggtct----ggaaa---ctttcctcgatccccgtcagctcggc
SEQ ID NO:3     375 tgatattggcat----cggga---catttgtcgatccacgccagcaaggc
SEQ ID NO:4     333 gtacgtcggtgaaaacggaga---atttgctcga----caatatcttagc
SEQ ID NO:5     126 --acttcaatca----gaaaatggctttcttggtttaactgca-------

SEQ ID NO:1     424 ggcaagctcaatgacgtaacca------aagaagacctcgtcaaactgat
SEQ ID NO:3     418 ggcaaactgaatgaagtcacta------aagaagacctgattaaactggt
SEQ ID NO:4     376 ggagagctcgagttggaattcacaccacaaggaacactcgccgaacgaat
SEQ ID NO:5     163 ---------tttgac-----cca------gaaaatgctaattcaaact---

SEQ ID NO:1     468 cgaagtcgatggtca---tgaacagcttttctacccgacc-----------
SEQ ID NO:3     462 cgagtttgataacaa---agaatatctctattacaaagcg-----------
SEQ ID NO:4     426 tcgtgcagctggtgccggtgttcccgcattctacac-accaacaggatac
SEQ ID NO:5     191 --tagtaaatgctgg---tggtcagcctt--------------------
```

```
SEQ ID NO:1   505 --ttcccgg--tcaacgtagctttcctccgcggtacgtatgctga---tg
SEQ ID NO:3   499 --attgcgc--cagatattgccttcattcgcgctaccacctgcga---ca
SEQ ID NO:4   475 ggtacccagattcaagaaggaggtgctccga-ttaagtacagtaaaactg
SEQ ID NO:5   215 ----------------------------gtggaa------ttaa---aa

SEQ ID NO:1   548 aatccggcaatatc-accatggacg--------aagaaatcgggcctttc
SEQ ID NO:3   542 gtgaaggctacgcc-acttttgaag--------atgaggtgatgtatctc
SEQ ID NO:4   524 aaaaaggaaagattgaagttgcaagtaaagcgaaagaaacacgacaattc
SEQ ID NO:5   227 aaggcggctcta---------------------------------ctttt

SEQ ID NO:1   589 ga---aagcacttccgta---gcccaggccgttcac--aactgtggcggt
SEQ ID NO:3   583 ga------cgcattggttattgcccaggcggtgcac--aataacggcggt
SEQ ID NO:4   574 aatggaattaattatgtaatggaagaggctatttggggagattttgcatt
SEQ ID NO:5   244 ga---tagtgctt---------------------t--ttctttcgcttt

SEQ ID NO:1   631 aaagtcgtcgtccaggtcaaagacgtcgtcgc-----tcacggcagcctc
SEQ ID NO:3   625 attgtgatgatgcaggtgcagaaaatggttaa-----gaaagccacgctg
SEQ ID NO:4   624 gatcaaggcgtggagagcagatac-tcttggaaatattcaattcagacat
SEQ ID NO:5   267 aa-----------------------------------------------ttc

SEQ ID NO:1   676 gacccgcgcatggtcaagatccctg--------gcatctatgtcgactac
SEQ ID NO:3   670 catcctaaatctgtccgtattccgg--------g---ttatctggtggat
SEQ ID NO:4   673 gctgctggaaatttcaataatccaatgtgcaaagcctctaaatgcac--c
SEQ ID NO:5   272 gtggcggtcatgtt---gatgcctg--------tgtgctaggtggact--

SEQ ID NO:1   718 gtcgtcgtagcagctccggaagaccatcagcag--acgtatgactgcgaa
SEQ ID NO:3   709 attgtggtggtcgatccg---gatcaaacccaa--ctgtatggcggtgca
SEQ ID NO:4   721 atcgtcgaagtag---aggaaatcgtcgaaccgggagtaattgctccaaa
SEQ ID NO:5   309 ------------------------------------------------

SEQ ID NO:1   766 t-------acgatccgtccctcagcggtgaacatcgtgctcctg-aaggc
SEQ ID NO:3   754 c-------cggttaaccgctttatttctggtgacttcacccttg-atgac
SEQ ID NO:4   768 cgatgtgcacattccatcaatctattgtcatcgtctagttttgggaaaga
SEQ ID NO:5   309 ----------------------------------------tg-aagtt

SEQ ID NO:1   808 gctac-------cgatgcagc---------tctccccatgagcgctaaga
SEQ ID NO:3   796 agtac-------caaacttag---------cctgcccctaaac-caacgt
SEQ ID NO:4   818 actacaaaaaaccaatcgaacggccaatgttcgcacacgaaggaccaata
SEQ ID NO:5   316 gatca-------agaagcaaa---------tctcgc--------------

SEQ ID NO:1   842 aaatcatcggc-cgccgcggcgctttggaattgactgaaaacgctgtcgt
SEQ ID NO:3   829 aaattagttgcgcggcgcgcattattcgaaatgcgtaaaggcgcggtggg
SEQ ID NO:4   868 aaaccatctac-atcggc--tgctggaaaatcgagagaaatcattg-cag
SEQ ID NO:5   336 -------------------------------taactgga-----------

SEQ ID NO:1   891 caacctcggcgtcggtgctcc-----ggaat--acgttgcttctgttgcc
SEQ ID NO:3   879 gaatgtcggcgtcggtattgc-----tgacg--gcattggcctggtcgcc
SEQ ID NO:4   914 cacgtgcagctttggagttcacagatggaatgtacgccaatttgggtatc
SEQ ID NO:5   344 ----------------------------------------tggtgcc

SEQ ID NO:1   934 gg--tgaagaaggtatcgccga------tacca-------ttaccctgac
SEQ ID NO:3   922 cg--agaagaaggttgtgctga------tgact-------ttattctgac
SEQ ID NO:4   964 gggattccgactttggcgccaaattatataccaaatggatttactgttca
SEQ ID NO:5   351 tg--gcaaaatggta-----------------------------------

SEQ ID NO:1   969 cgtcgaaggtg------gcgccatcggtggcgt-accgcagggcggtgcc
SEQ ID NO:3   957 ggtagaaacag------gtccgattggcggaattacttcacaggggatcg
SEQ ID NO:4  1014 tttgcaaagtgagaatggtattattggagtggg-accata------tcca
SEQ ID NO:5   364 ------------------------------------------------
```

```
SEQ ID NO:1    1012 cgcttcggttcgtcccgca-atgccgatgccatca----tcgaccacacc
SEQ ID NO:3    1001 c-ctttggcgcgaacgtga-atacccgtgccattc----tggatatgacg
SEQ ID NO:4    1057 agaaaag----gaacagaagacgccgatctcattaatgctggaaaagagc
SEQ ID NO:5     364 -------------ccagga-atg--------------------------

SEQ ID NO:1    1057 tatcagttcgacttctacgatggcggc----------ggtctggacatcg
SEQ ID NO:3    1045 tcccagtttgatttttatcacggtggc----------ggtctggatgttt
SEQ ID NO:4    1103 ---caattactcttct-caaaggagcttcaattgttggttctgatgaatc
SEQ ID NO:5     373 ---------------------ggcgga----------gcaatggacttag

SEQ ID NO:1    1097 cttacctcggcctgg-----cccagtgcgatg-------gctcgggcaac
SEQ ID NO:3    1085 gttatttgagttttg-----ctgaagtcgacc-------agcacggtaac
SEQ ID NO:4    1149 attcgcaatgattcgtggttctcatatggatattactgtgctcggtgcac
SEQ ID NO:5     392 ------------------------------tg-------actggtgcaa-

SEQ ID NO:1    1135 atcaacgtcagca-agttcggtactaacgttgccggctgcggcggtttcc
SEQ ID NO:3    1123 gtcggcgtgcata-aattcaatggtaaaatcatgggcaccggtggattta
SEQ ID NO:4    1199 ttca--gtgctcacagtttgg----agatttagcgaattggatgattccg
SEQ ID NO:5     404 --------------------------------------------------

SEQ ID NO:1    1184 ccaacatt--tcccagcagacaccgaatgtttacttctgcggcacct-tc
SEQ ID NO:3    1172 ttgatatcagtgccacttcgaagaaaatcatt--ttctgcggcacat-ta
SEQ ID NO:4    1243 ggaaaatt-------ggtga-aaggaatgggcggtgcaatggatcttgtc
SEQ ID NO:5     404 ---------------------aaaaagtgattatt-----ggca------

SEQ ID NO:1    1231 acggctggcggcttgaaaatcgctgtcgaagacggcaaagtcaagatcct
SEQ ID NO:3    1219 actgcgggcagtttaaaaacagaaattaccgacggcaaattaaatatcgt
SEQ ID NO:4    1285 tctgctcccgg------agcccgtgt-gatcgttgtaatggagcatgtat
SEQ ID NO:5     422 ------------tggaacattg-----------tgccaagtcaggttcct

SEQ ID NO:1    1281 ccaggaaggcaaagccaagaagttcatcaaagctgtcgaccagatcactt
SEQ ID NO:3    1269 ccaggaaggacgggtgaagaaatttattcgggaactaccggaaattactt
SEQ ID NO:4    1328 cgaagaacggagagccaaaaatt---------------ctagagcactg
SEQ ID NO:5     449 caaaaattctaaag---aaatgtacattaccgct------cacagcaagt

SEQ ID NO:1    1331 tcaacgg-----ttcctatgcagc---ccgcaacggcaaacacgttctct
SEQ ID NO:3    1319 tcagcggaaaaatcgctctcgagc---gagggctgg-----atgttcgtt
SEQ ID NO:4    1362 cgaac-------ttcctctga--c---cggcaaagg--agtaatttcccg
SEQ ID NO:5     490 aaaaaag-----ttgccatggtggttaccgaattggca-----gtattta

SEQ ID NO:1    1373 a--catcacagaacgctgcgtatttgaactgacca--aagaa-ggcttga
SEQ ID NO:3    1361 a--tatcactgagcgcgcagtattcacgctgaaag--aagac-ggcctgc
SEQ ID NO:4    1398 aatcattactgatatggcagttttcgacgtggacacaaagaacggattga
SEQ ID NO:5     530 a--cttcattgaaggcagattagttcta------a--aagaa---catgc

SEQ ID NO:1    1418 aactcatcgaagtcgcaccgggcatcgatattgaaaaagatatcctcgct
SEQ ID NO:3    1406 atttaatcgaaatcgcccctggcgtcgatttacaaaaagatattctcgac
SEQ ID NO:4    1448 cattgatcgaagt--caggaaggatc-ttactgtagatgatat-------
SEQ ID NO:5     567 tcctcat----------------gtggatttagaaaca---attaaagcc

SEQ ID NO:1    1468 cacatggacttcaagccgat--cattgata---atccga--aactcatgg
SEQ ID NO:3    1456 aaaatggatttcaccccagt--gatttcgccagaactca--aactgatgg
SEQ ID NO:4    1488 --caagaaactca--ccg-----cttgcaa---attcga--aatttccga
SEQ ID NO:5     598 aaaacag-----aagccgatttcattgtt-----gccgatgatttcaaag

SEQ ID NO:1    1511 atgcccgcctcttccaggacggtcccatggga------ctgaaaaaa---
SEQ ID NO:3    1502 acgaaagattatttatcgatgcggcgatgggttttgtcctgcctgaagcg
SEQ ID NO:4    1524 aaatctgaagccaatgggacaggctcctctta------atcaaggataa-
SEQ ID NO:5     638 aaatgcaaatcagccag---------aaagga------cttgaattatga
```

```
SEQ ID NO:1     1552 ------taa
SEQ ID NO:3     1552 gctcattaa
SEQ ID NO:4     1567 ---------
SEQ ID NO:5      673 ---------
```

**Figure 9**

```
SEQ ID NO:2      1 ------------------------mrkveiit---------aeqaaqlv
SEQ ID NO:6      1 --------------------------mpvls---------aqeavnyi
SEQ ID NO:7      1 mpilskiwaapaagilrktprnahqmrlismtssmkakvfnsaeeavkdi
SEQ ID NO:8      1 ------------------------mnakeli----------arriamel

SEQ ID NO:2     17 kdndtitsigfvssahpealt--kalekrfldtntpqnltyiyagsqgkr
SEQ ID NO:6     14 pdeatlcvlg-agggileattlitaladkykqtqtprnlsiisptglgdr
SEQ ID NO:7     51 pdnakllvggfglcgipenli--qai------tktgqkgltcvsnnagv-
SEQ ID NO:8     16 hdgd-ivnlg-------------------------------------

SEQ ID NO:2     65 dgraaehlahtgllkraiighwqtvpaigklavenkieaynfsqgtlvhw
SEQ ID NO:6     63 adrgisplaqeglvkwalcghwgqspriselaeqnkiiaynypqgvltqt
SEQ ID NO:7     92 dnwglglllqtrqikkmissyvgengefarqylsgeleleftpqgtlaer
SEQ ID NO:8     25 --------------------------------------------------

SEQ ID NO:2    115 fralaghklgvftdigletfldprqlggklndvtkedlvkliev------
SEQ ID NO:6    113 lraaaahqpgiisdigigtfvdprqqggklnevtkedliklvef------
SEQ ID NO:7    142 iraagagvpafytptgygtqi---qeggapikysktekgk-ievaskake
SEQ ID NO:8     25 -------------igl---------------------------------

SEQ ID NO:2    159 ----dgheqlfyptfpvnvaflrgtyadesgnitmdeeigpfestsvaqa
SEQ ID NO:6    157 ----dnkeylyykaiapdiafirattcdsegyatfedevmyldalviaqa
SEQ ID NO:7    188 trqfnginyvmeeaiwgdfalikawradtlgniqfrhaagnfnnpmckas
SEQ ID NO:8     28 ------------ptqvvn-------ylpdnvnitlqsengflglta----

SEQ ID NO:2    205 vhncggkvvvqvkdvvahgsldprmvkipgiyvdyvvvaapedhqqtydc
SEQ ID NO:6    203 vhnnggivmmqvqkmvkkatlhpksvripgylvd-ivvvdpdqtqlygga
SEQ ID NO:7    238 --kc---tiveveeivepgviapndvhipsiychrlvlg------knykk
SEQ ID NO:8     55 --------------------------------------------------

SEQ ID NO:2    255 eydpslsgehrapegatdaalpmsakkiigrrgaleltenavvnlgvg--
SEQ ID NO:6    252 pvnrfisgdftl-ddstklslplnqrklvarralfemrkgavgnvgvg--
SEQ ID NO:7    277 pierpmfahegpikpstsaa--gksreiiaaraaleftdgmyanlgigip
SEQ ID NO:8     55 -fdp-----------------------------------enansnl-vn--

SEQ ID NO:2    303 --apeyvasvageegiadtitltveggaig--gvpqggarfgssrnad--
SEQ ID NO:6    299 --iadgiglvareegcaddfiltvetgpig--gitsqgiafganvntr--
SEQ ID NO:7    325 tlapnyipn-------gftvhlqsengiigvgpyprkgtedadlinagke
SEQ ID NO:8     67 --a----------------------ggqpc--gikkggstf---------

SEQ ID NO:2    347 --------aiidhtyqfdfydgggldiaylglaqcdgsgni-nvskfgtn
SEQ ID NO:6    343 --------aildmtsqfdfyhgggldvcylsfaevdqhgnv-gvhkfngk
SEQ ID NO:7    368 pitllkgasivgsdesfamirgshmditvlgalqcsqfgdlanwmipgkl
SEQ ID NO:8     82 -----------dsafsfalirgghvdacvlgglevdqeanlanwmvpgkm

SEQ ID NO:2    388 vagcggfpnisqqtpnvyfcgtftagglkiav------edgkvkilqegk
SEQ ID NO:6    384 imgtggfidisatskkiifcgtltagslktei------tdgklnivqegr
SEQ ID NO:7    418 vkgmggamdl------------vsapgarvivvmehvskngepkilehce
SEQ ID NO:8    121 vpgmggamdlvtgakkvii-------gmehca------ksgsskilk---

SEQ ID NO:2    432 akkfikavdqitfngsyaarngkhvl--yitercvfel-tkeglklieva
SEQ ID NO:6    428 vkkfirelpeitfsgkialergldvr--yiteravftl-kedglhlieia
SEQ ID NO:7    456 -----------------lpltgkgvisriitdmavfdvdtkngltlievr
SEQ ID NO:8    155 -----kctlplt-------askkvam--vvtelavfnf-iegrlvlkeha
```

```
SEQ ID NO:2     479 pgidiekdi--lahmdfkpiidnp-klmdarlfqdgpmglkk-----
SEQ ID NO:6     475 pgvdlqkdi--ldkmdftpvispelklmderlfidaamgfvlpeaah
SEQ ID NO:7     489 kdltvd-dikkltackfe-isenl-kpmgqaplnqg----------
SEQ ID NO:8     190 phvdle-ti--kakteadfivad-----dfkemqisqkglel-----
```

**Figure 10**

GTGAAAACTGTGTATACTCTCGGAATCGACGTTGGTTCTTCTTCTTCCAAGGCAGTCATC
CTGGAAGATGGCAAGAAGATCGTCGCCCATGCCGTCGTTGAAATCGGCACCGGTTCGACC
GGTCCGGAACGCGTCCTGGACGAAGTCTTCAAAGATACCAACTTAAAAATTGAAGACATG
GCGAACATCATCGCCACAGGCTATGGCCGTTTCAATGTCGACTGCGCCAAAGGCGAAGTC
AGCGAAATCACGTGCCATGCCAAAGGGGCCCTCTTTGAATGCCCCGGTACGACGACCATC
CTCGATATCGGCGGTCAGGACGTCAAGTCCATCAAATTGAATGGCCAGGGCCTGGTCATG
CAGTTTGCCATGAACGACAAATGCGCCGCTGGTACGGGCCGTTTCCTCGACGTCATGTCG
AAGGTACTGGAAATCCCCATGTCTGAAATGGGGGACTGGTACTTCAAATCGAAGCATCCC
GCTGCCGTCAGCAGTACCTGCACGGTTTTTGCTGAATCGGAAGTCATTTCCCTTCTTTCC
AAGAATGTCCCGAAAGAAGATATCGTAGCCGGTGTCCATCAGTCCATCGCCGCCAAAGCC
TGCGCTCTCGTGCGCCGCGTCGGTGTCGGTGAAGACCTGACCATGACCGGCGGTGGCTCC
CGCGATCCCGGCGTCGTCGATGCCGTATCGAAAGAATTAGGTATTCCTGTCAGAGTCGCT
CTGCATCCCCAAGCGGTGGGTGCTCTCGGAGCTGCTTTGATTGCTTATGATAAAATCAAG
AAATAA (SEQ ID NO:9)

**Figure 11**

VKTVYTLGIDVGSSSSKAVILEDGKKIVAHAVVEIGTGSTGPERVLDEVFKDTNLKIEDM
ANIIATGYGRFNVDCAKGEVSEITCHAKGALFECPGTTTILDIGGQDVKSIKLNGQGLVM
QFAMNDKCAAGTGRFLDVMSKVLEIPMSEMGDWYFKSKHPAAVSSTCTVFAESEVISLLS
KNVPKEDIVAGVHQSIAAKACALVRRVGVGEDLTMTGGGSRDPGVVDAVSKELGIPVRVA
LHPQAVGALGAALIAYDKIKK (SEQ ID NO:10)

**Figure 12**

```
SEQ ID NO:9       1 gtgaaaactgtgtatactctcggaatcgacgttggttcttcttcttccaa
SEQ ID NO:11      1 ---atgagtatctataccttgggaatcgatgttggatctactgcatccaa
SEQ ID NO:12      1 gtggcagtggcatattcgattggcattgattccggctcaaccgccaccaa
SEQ ID NO:13      1 ------------atgattttagggatagatgttggatctacaacaacgaa

SEQ ID NO:9      51 ggcagtcatcctggaagatggcaagaagatcgtcgc-ccatgccgtcgtt
SEQ ID NO:11     48 gtgcattatcctgaaagatggaaaagaaatcgtggc-gaaatccctggta
SEQ ID NO:12     51 agggatcttactggcagacggcgtgatta----cgcgccgtttcctcgtt
SEQ ID NO:13     39 gatggttctaatggaagatagc---aagataatttg-gtataagatagag

SEQ ID NO:9     100 gaaatcggcaccggttcgaccggtccggaacgcgtcctggacgaagtctt
SEQ ID NO:11     97 gccgtggggaccggaacttccggtcccgcacggtctatttcggaagtcct
SEQ ID NO:12     97 ccaa----ccccctttcgcccgg-caacagcaattact----gaagcctg
SEQ ID NO:13     85 gatattgg-agttgtta-----ttgaggaagatattttattaaaaatggt

SEQ ID NO:9     150 caaagatacc-aacttaaaaattgaagacatggcgaacatcatcgc-cac
SEQ ID NO:11    147 ggaaaatgcc-cacatgaaaaaagaagacatggcctttaccctggc-tac
SEQ ID NO:12    138 ggaa-actct-gcgcgaagggttagagacaacgccgtttctgacgctcac
SEQ ID NO:13    129 taaggagattgaacaaaaatatccaatagat----aaaatcgttgc-aac

SEQ ID NO:9     198 aggctatggccgtttcaatgtcg---------actgcgccaaaggcgaag
SEQ ID NO:11    195 cggctacggacg---caat-tcgctggaaggcattgccgacaagcaga--
SEQ ID NO:12    186 cggctacgggcggcaactggtgg---------attttgccgataaacagg
SEQ ID NO:13    174 tggatatggaaggcataaggtta---------gttttgcagataagatag

SEQ ID NO:9     239 tcagcgaaatcacgtgccatgccaaaggggcc---ctctttgaatgcccc
SEQ ID NO:11    239 tgagcgaactgagctgccatgccatgggcgcc---agctttatctggccc
SEQ ID NO:12    227 taacggaaatctcctgtcacgggctgggcgca---cggtttcttgcgcca
SEQ ID NO:13    215 ttccagaagtta-ttgcattgggaaaaggagctaactatttctttaacga

SEQ ID NO:9     286 ggtacgacga--ccatcctcgatatcggcggtcaggacgtcaa-gtccat
SEQ ID NO:11    286 --aacgtccataccgtcatcgatatcggcgggcaggatgtgaa-ggtcat
SEQ ID NO:12    274 gcaacgcgcg--cggtaatcgacatcggtggtcaggacagcaaagtgatt
SEQ ID NO:13    264 ggcagatgga----gttatagacattggagggcaagatacaaa-ggtctt

SEQ ID NO:9     333 caaattga--atggccagggcctggtcatgcagtttgcc-atgaacgaca
SEQ ID NO:11    333 ccatgtgg--aaaacgggaccatgacca---atttccag-atgaatgata
SEQ ID NO:12    322 cagcttgatgatgacggtaacctg----tgcgatttcctgatgaatgaca
SEQ ID NO:13    309 aaagattg--ataaaaacggaaaagttgttgattttatc-ctatcagata

SEQ ID NO:9     380 aatgcgccgctggtacgggccgtttcctcgacgtcatgtcgaaggtactg
SEQ ID NO:11    377 aatgcgctgccgggactggccgtttcctggatgttatggccaatatcctg
SEQ ID NO:12    368 aatgcgcggcgggcaccgggcgtttcctggaggtgatctcgcgcacgctt
SEQ ID NO:13    356 aatgtgccgctggaactggaaaattcttaga---------aaaggcatta

SEQ ID NO:9     430 gaaatccccatgtct-ga--aatggggactggtactt-caaatcgaagc
SEQ ID NO:11    427 gaagtgaaggtttcc-ga--cctggctgagctgggagc-caaatccacca
SEQ ID NO:12    418 ggca--ccagcgtcgagc--aactcgacagcattaccg-aaaat---gtc
SEQ ID NO:13    397 gatattttaaaaatt-gataaaaatgagataaataaatacaaatcagata

SEQ ID NO:9     476 atcccgct-gccgtcagcagtacctgcacggtttttgctgaatcggaagt
SEQ ID NO:11    473 aacgggtg-gctatcagctccacctgtactgtgtttgcagaaagtgaagt
SEQ ID NO:12    460 acgccgcacgccatcacgagtatgtgcacagtgtttgctgaatcagaagc
SEQ ID NO:13    446 atatcgct-aaaatatcttcaatgtgtgctgtctttgctgaaagtgagat
```

```
SEQ ID NO:9     525 catttcccttctttccaagaatgtcccgaaagaa--gatatcgtagccgg
SEQ ID NO:11    522 catcagccagctgtccaa--aggaaccgacaagatcgacatcattgccgg
SEQ ID NO:12    510 gatcagcctgcgctcagcgggcgtcgcgccagaa--gcgattctcgcagg
SEQ ID NO:13    495 aataagcttactatcaaaaaaagttccaaaggaa--ggcattttaatggg

SEQ ID NO:9     573 tgtccatcagtccatcgccgccaaagcctgcgctctcgtgc-gccgcgtc
SEQ ID NO:11    570 gatccatcgttctgtagccagccgggtcattggtcttgcca-atcgggtg
SEQ ID NO:12    558 agtgattaacgcgat-ggcgcggaggagtgc-caatttcat-tgctcgtc
SEQ ID NO:13    543 cgtctatgagagtat------aataaataggattatcccaatgaccaata

SEQ ID NO:9     622 ggtgtcgg--tgaagacctgaccatgaccggcggtggctcccgcgat--c
SEQ ID NO:11    619 gggattgt--gaaagacgtggtcatgaccggcggtgtagcccagaac--t
SEQ ID NO:12    605 tctc-ctg--tgaagcgccgattctgtttactggtggcgttagtcattgc
SEQ ID NO:13    587 ggcttaaaattcaaaacatagtgtttagtggaggagttgctaaaaat--a

SEQ ID NO:9     668 ccggcgtcgtcgatgccgtatcgaaagaat------taggtattcctgtc
SEQ ID NO:11    665 atggcgtgagaggagccct------ggaag------aaggccttggcgtg
SEQ ID NO:12    652 cagaagt-----ttgcccggatgctggaatctcacctgcgaatgccggta
SEQ ID NO:13    635 aggttttggttgagatgtttgagaaaaaat------tgaataaaaaacta

SEQ ID NO:9     712 agagtcgctctgcatccccaagcggtg------ggtgctctcggagctgc
SEQ ID NO:11    703 gaaatcaagacgtctcccctggctcagtacaacggtgccctgggtgccgc
SEQ ID NO:12    697 aatacccatcctgatgcgcaatttgct------ggcgcaattggcgcggc
SEQ ID NO:13    679 ctaattccaaaagaaccacagattgtt------tgctgtgttggagctat

SEQ ID NO:9     756 tttgattgctta-----tgataaaatcaagaaa-taa
SEQ ID NO:11    753 tctgtatgcgta-----t-aaaaaagcagccaaataa
SEQ ID NO:12    741 ggtaattggtcaacgagtgaggacacgccgatga---
SEQ ID NO:13    723 attggtt----------taa----------
```

**Figure 13**

```
SEQ ID NO:10      1 vktvytlgidvgsssskaviledgkkivahavveigtgstgpervldevf
SEQ ID NO:14      1 ms-iytlgidvgstaskciilkdgkeivakslvavgtgtsgparsisevl
SEQ ID NO:15      1 mavaysigidsgstatkgilladg-vitrrflvpt---pfrpataiteaw
SEQ ID NO:16      1 ----milgidvgstttkmvlmeds-kiiwykiedigv--vieedillkmv

SEQ ID NO:10     51 kdtnlkiedmaniiatgygrfnvd-cakgevseitchakgalfecpgttt
SEQ ID NO:14     50 enahmkkedmaftlatgygrnslegiadkqmselschamgasfiwpnvht
SEQ ID NO:15     47 etlreglettpfltltgygrqlvd-fadkqvteischglgarflapatra
SEQ ID NO:16     44 keieqkyp-idkivatgygrhkvs-fadkivpevialgkganyffneadg

SEQ ID NO:10    100 ildiggqdvksiklngqglvmqfamndkcaagtgrfldvmskvleipmse
SEQ ID NO:14    100 vidiggqdvkvihve-ngtmtnfqmndkcaagtgrfldvmanilevkvsd
SEQ ID NO:15     96 vidiggqdskviqlddd gnlcdflmndkcaagtgrflevisrtlgtsveq
SEQ ID NO:16     92 vidiggqdtkvlkidkngkvvdfilsdkcaagtgkflekaldilkidkne

SEQ ID NO:10    150 mgdwyfkskhpaavsstctvfaesevisllsknvpkedivagvhqsiaak
SEQ ID NO:14    149 laelgakstkrvaisstctvfaesevisqlskgtdkidiiagihrsvasr
SEQ ID NO:15    146 l-dsitenvtphaitsmctvfaeseaislrsagvapeailagvinamarr
SEQ ID NO:16    142 ink--yksdniakissmcavfaeseiisllskkvpkegilmgvyesiinr

SEQ ID NO:10    200 acalvrrvgvgedltmtgggsrdpgvvdavskelgipvrvalhpqavgal
SEQ ID NO:14    199 viglanrvgivkdvvmtggvaqnygvrgaleeglgveiktsplaqyngal
SEQ ID NO:15    195 sanfiarlsceapilftggvshcqkfarmleshlrmpvnthpdaqfagai
SEQ ID NO:16    190 vipmtnrlki-qnivfsggvaknkvlvemfekklnkkllipkepqivccv

SEQ ID NO:10    250 gaaliaydkikk--
SEQ ID NO:14    249 gaalyaykkaak--
SEQ ID NO:15    245 gaavig-qrvrtrr
SEQ ID NO:16    239 gailv---------
```

**Figure 14**

ATGAGTGAAGAAAAAACAGTAGATATTGAAAGCATGAGCTCCAAGGAAGCCCTTGGTTAC
TTCTTGCCGAAAGTCGATGAAGACGCACGTAAAGCGAAAAAAGAAGGCCGCCTCGTTTGC
TGGTCCGCTTCTGTCGCTCCTCCGGAATTCTGCACGGCTATGGACATCGCCATCGTCTAT
CCGGAAACTCACGCAGCTGGTATCGGTGCCCGTCACGGTGCTCCGGCCATGCTCGAAGTT
GCTGAAAACAAAGGTTACAACCAGGACATCTGTTCCTACTGCCGCGTCAACATGGGCTAC
ATGGAACTCCTCAAACAGCAGGCTCTGACAGGCGAAACGCCGGAAGTCCTCAAAAACTCC
CCGGCTTCTCCGATTCCCCTTCCGGATGTTGTCCTCACTTGCAACAACATCTGCAATACC
TTGCTCAAATGGTATGAAAACTTGGCTAAAGAATTGAACGTACCTCTCATCAACATCGAC
GTACCGTTCAACCATGAATTCCCTGTTACGAAACACGCTAAACAGTACATCGTCGGCGAA
TTCAAACATGCTATCAAACAGCTCGAAGACCTTTGCGGCCGTCCCTTCGACTATGACAAA
TTCTTCGAAGTACAGAAACAGACACAGCGCTCCATCGCTGCCTGGAACAAAATCGCTACG
TACTTCCAGTACAAACCGTCGCCGCTCAACGGCTTCGACCTCTTCAACTACATGGGCCTC
GCCGTTGCTGCCCGCTCCTTGAACTACTCGGAAATCACGTTCAACAAATTCCTCAAAGAA
TTGGACGAAAAAGTAGCTAATAAGAAATGGGCTTTCGGTGAAAACGAAAAATCCCGTGTT
ACTTGGGAAGGTATCGCTGTCTGGATCGCTCTCGGCCACACCTTCAAAGAACTCAAAGGT
CAGGGCGCTCTCATGACTGGTTCCGCTTATCCTGGCATGTGGGACGTTTCCTACGAACCG
GGCGACCTCGAATCCATGGCAGAAGCTTATTCCCGTACATACATCAACTGCTGCCTCGAA
CAGCGCGGTGCTGTTCTTGAAAAAGTTGTCCGCGATGGCAAATGCGACGGCTTGATCATG
CACCAGAACCGTTCCTGCAAGAACATGAGCCTCCTCAACAACGAAGGCGGCCAGCGCATC
CAGAAGAACCTCGGCGTACCGTACGTCATCTTCGACGGCGACCAGACCGATGCTCGTAAC
TTCTCGGAAGCACAGTTCGATACCCGCGTAGAAGCTTTGGCAGAAATGATGGCAGACAAA
AAAGCCAATGAAGGAGGAAACCACTAA (SEQ ID NO:17)

**Figure 15**

MSEEKTVDIESMSSKEALGYFLPKVDEDARKAKKEGRLVCWSASVAPPEFCTAMDIAIVY
PETHAAGIGARHGAPAMLEVAENKGYNQDICSYCRVNMGYMELLKQQALTGETPEVLKNS
PASPIPLPDVVLTCNNICNTLLKWYENLAKELNVPLINIDVPFNHEFPVTKHAKQYIVGE
FKHAIKQLEDLCGRPFDYDKFFEVQKQTQRSIAAWNKIATYFQYKPSPLNGFDLFNYMGL
AVAARSLNYSEITFNKFLKELDEKVANKKWAFGENEKSRVTWEGIAVWIALGHTFKELKG
QGALMTGSAYPGMWDVSYEPGDLESMAEAYSRTYINCCLEQRGAVLEKVVRDGKCDGLIM
HQNRSCKNMSLLNNEGGQRIQKNLGVPYVIFDGDQTDARNFSEAQFDTRVEALAEMMADK
KANEGGNH (SEQ ID NO:18)

Figure 16

```
SEQ ID NO:17      1 atgagtgaagaaaaaacagtagatattgaaagcatgagctccaaggaagc
SEQ ID NO:19      1 atg------ccaaagacagta---------agccctggcgttcagg----
SEQ ID NO:20      1 ----atgatgaaattaaag--gcaattgaaaagttga--tgcaa------
SEQ ID NO:21      1 ------------atgtcacttgtcaccga----------tcta--cccgc

SEQ ID NO:17     51 cctt---ggttacttcttgccgaaa--gtcgatgaagacgca-------c
SEQ ID NO:19     32 -cat---tgagagatgtagttgaaaaggtttacagagaactg-------c
SEQ ID NO:20     37 ----------aaatt-----cgcca--gtagaaaagaacagc-------t
SEQ ID NO:21     27 cattttcgatcagttct--ctgaag--ctcgccagacaggctttctcacc

SEQ ID NO:17     89 gta-aagcgaaaa-aagaaggccgcctcgttt-gctggtccgcttctgtc
SEQ ID NO:19     71 ggg-aaccgaaag-aaagaggagaaaaagtag-gctggtcctcttc--ca
SEQ ID NO:20     63 atataagcaaaaagaagaaggtagaaaagttt---ttggaatgttctgtg
SEQ ID NO:21     73 gtc-atggatctc-aaggag--cgcggcattccgctggt-------tggc

SEQ ID NO:17    136 gctcctccggaattctgcacggctatggacatcgccatcgtc--tatccg
SEQ ID NO:19    116 agttcccctgcgaactggctgaatcttttcggctgcatgttgggtatccg
SEQ ID NO:20    110 cct--------atgttcca-----atagaaat---aat--tt--tagcag
SEQ ID NO:21    112 act---------tactgcacctttatg----ccgcaagag-----atccc

SEQ ID NO:17    184 gaaactca--cgcagctggtatcggtgcc---cgtcacggtg--------
SEQ ID NO:19    166 gaaaacca--ggctgctggtatcgctgccaaccgtgacggcgaagtgatg
SEQ ID NO:20    140 caaatgcaatcccagttggtttgtgtgga---ggtaaaaat---------
SEQ ID NO:21    144 ga-----t--ggcagc------cggtgcg---gtt----gtg--------

SEQ ID NO:17    221 ----------------------------------------ctccggccatgc
SEQ ID NO:19    214 tgccaggctgcagaagatatcggttatgacaacgatatctgcggctatgc
SEQ ID NO:20    178 ---------------------------------------gacacaa
SEQ ID NO:21    166 ----------------------------------------gtttcgctctgt

SEQ ID NO:17    233 tcgaagt-t----------gctg--------------------aaaa--
SEQ ID NO:19    264 ccgtatt-tccctggcttatgctgccgggttccggggtgccaacaaaatg
SEQ ID NO:20    185 tcccaat-a----------gcag--------------------a-----
SEQ ID NO:21    178 tccacctct----------gatg--------------------aaac--

SEQ ID NO:17    249 --caaaggttacaaccaggacatctgttcctactgccgcgtcaacatg--
SEQ ID NO:19    313 gacaaagatggcaactatgtcatcaacccccacagcggcaaacagatgaa
SEQ ID NO:20    198 ---ggaggat-ttgccaagaaacctatgcc--------cattaata----
SEQ ID NO:21    195 --ca---ttgaagaagcggagaaagat-----ctgccgcg-caacct---

SEQ ID NO:17    295 ----------ggctacatggaactc--ctcaaacagcag-----------
SEQ ID NO:19    363 agatgccaatggcaaaaaggtattcgacgcagatggcaaacccgtaatcg
SEQ ID NO:20    232 --------------------aaatc--atccta----tg-----------
SEQ ID NO:21    231 ------------ctgcccg---ctg--attaaa-agca------------

SEQ ID NO:17    322 ------------------------------------------------
SEQ ID NO:19    413 atcccaagaccctgaaacccttgccaccaccgacaacatctatgaaatc
SEQ ID NO:20    245 ------------------------------------------------
SEQ ID NO:21    251 ------------------------------------------------

SEQ ID NO:17    322 ---gctctgac---aggcgaaa-----------cgccggaa-gtcctcaa
SEQ ID NO:19    463 gctgctctgccggaaggggaagaaaagacccgccgccagaatgccctgca
SEQ ID NO:20    245 ---gttttaa-----------------------------gaa-ggca--aa
SEQ ID NO:21    251 ---gctacggc---ttcggcaa-----------aaccg----------at
```

```
SEQ ID NO:17    354 aaactccccggcttctccgattccccttccggatgttgtcctcacttgca
SEQ ID NO:19    513 caaatatcgtcagatgaccatgcccatgccggacttcgtgctgtgctgca
SEQ ID NO:20    261 aacctgccc--ttactttgaagcatct----gatatagttat-tggagaa
SEQ ID NO:21    274 aaatgcccctacttct-------acttttcggatctggtggtc---ggtg

SEQ ID NO:17    404 acaacatctgca-------ataccttgctcaaatggtatgaaaacttgg-
SEQ ID NO:19    563 acaacatctgca-------actgcatgaccaaatggtatgaagacattg-
SEQ ID NO:20    304 actacctgtgaaggaaagaagaagatgtttgagttgatggagagattggt
SEQ ID NO:21    314 aaaccacctgcg-------acggcaaaaagaaaatgtatgaatacatgg-

SEQ ID NO:17    446 -ctaaagaattgaac---gtacctctca----tcaacatcgacgtac--c
SEQ ID NO:19    605 -cccgtcggcacaac---attcctttga----tcatgatcgacgttc--c
SEQ ID NO:20    354 gccaatgcatataat---gcacctcccacacatgaaagatgaagatt--c
SEQ ID NO:21    356 -c---ggagtttaagcctgttcatgtga----tgca-attgcccaacagc

SEQ ID NO:17    486 gttca--accatgaattc---cctg--tta-cgaa----ac--acgctaa
SEQ ID NO:19    645 ttaca--ac---gaattcgaccatg--tcaacgaa----gccaacgtgaa
SEQ ID NO:20    399 tttga--a-----aatct---ggat--taa-agaagttgaa--aagctaa
SEQ ID NO:21    397 gttaaggacgatgcctcg---cgtgcgtta-tgga----a----------

SEQ ID NO:17    522 acagtacatcgtcg-----gcgaattcaaacatgctatca----aacagc
SEQ ID NO:19    684 a---tacatccggt-----cccagctggatacggccatcc----gtcaaa
SEQ ID NO:20    434 --aagaattggttgagaaagagactggaaataaaataacagaggaaaagt
SEQ ID NO:21    429 --agccgagatgct-----gcgcttgcaa----------a----aaacgg

SEQ ID NO:17    563 tcgaagacctttgcggccgtcccttcgactatgacaaattcttcgaagta
SEQ ID NO:19    722 tggaagaaatcaccggcaagaagttcgatgaagacaaattc------gaa
SEQ ID NO:20    482 taaaaga------------------------gacagttgat--aaagta
SEQ ID NO:21    458 tagaagaacgttttgggcacgagattagcgaagatgctctgcgcgatgcc

SEQ ID NO:17    613 cagaaacagacacagcgctc-catcg--ctgcc------tggaacaaaat
SEQ ID NO:19    766 cag-tgctgccagaacgc-c-aaccgtactgccaaagcatggctgaaggt
SEQ ID NO:20    505 aataaagttagggag----------------t------tgttttataaa
SEQ ID NO:21    508 attgcgctgaaaaaccgcgaacgtcg--cgcac------tgg---ctaat

SEQ ID NO:17    654 cgctacgtacttc--c--agtacaaaccgtcgccgctcaacggcttcgac
SEQ ID NO:19    813 ttgcgactacctg--c--agtacaaaccggctccgttcaacgggttcgac
SEQ ID NO:20    532 ctctatgaattga--ggaagaataaaccagctccaattaagggtttagat
SEQ ID NO:21    547 ttttatcatcttgggc--agttaaatcctccggcgcttagcggcagcgac

SEQ ID NO:17    700 ctcttcaactacatgggcctcgccg-ttgctgcccgctccttgaactact
SEQ ID NO:19    859 ctgttcaaccatatggctgacgtgg-ttaccgcccgtggccgtgtggaag
SEQ ID NO:20    580 gttttaaaattattccagtttgcctatttattggatattgatgacacaat
SEQ ID NO:21    595 attctga---aagtggtttacggcg-caaccttccggttcgataaagagg

SEQ ID NO:17    749 cggaaatcacgttcaacaaattcctcaaagaattggacgaaaaagtagc-
SEQ ID NO:19    908 ctgctgaagctttcgaactgctggccaaggaactggaacagcatgt----
SEQ ID NO:20    630 agggatt----ttagaggatttaattgaggagttagaggagagagtt---
SEQ ID NO:21    641 cg------------------ttgatcaatgaactggatgcaatgaccgcc

SEQ ID NO:17    798 -----taataagaaatgggctttcggtgaa------aacgaaaaatcccg
SEQ ID NO:19    954 --------gaaggaaggcaccaccaccgctcccttcaaagaacagcatcg
SEQ ID NO:20    673 -----aaaaaaggagaaggttatgaaggaa------agagaa--------
SEQ ID NO:21    673 cgcgttcgtcagcagtgggaagaag--gcc------agcgactggacccg

SEQ ID NO:17    837 tgttacttgggaaggta-tcgctgtctggatcgctctcggccacacc---
SEQ ID NO:19    996 tatcatgttcgaaggga-tcccctgctgg--ccgaaactgccgaacc---
SEQ ID NO:20    704 -------ttttaataac-tggctgtc-caatggttgctggaaacaataag
SEQ ID NO:21    715 cgt--ccgcgcattttaatcaccggctg---cccgattggcggcgc----
```

```
SEQ ID NO:17    883 ----t--tcaaagaactca--aaggtcagggcgctctcatgactggttcc
SEQ ID NO:19   1040 ----tgttcaaaccgctga--aagccaacggcctgaacatcaccggcgtt
SEQ ID NO:20    745 attgt--tgaaattattgaggaagtt---ggaggagtagttgttggtgaa
SEQ ID NO:21    756 ------------agcaga--aaaagtggtgcgcgcgattgaagagaatg

SEQ ID NO:17    925 gcttat---cctggcatgtgggacgtttcctacgaacc--------ggg-
SEQ ID NO:19   1084 gtatatgctcctgctttcgggttcgtgtacaacaacct--------gga-
SEQ ID NO:20    790 g--aaa---gctgcactggaacaagattctttgaaaactttgttgaggg-
SEQ ID NO:21    791 gc---g---gctgggttgtcggttatgaaaactgcacc--------gggg

SEQ ID NO:17    963 ------cga---cctcg-aatccatggcagaa----gcttattcccgtac
SEQ ID NO:19   1125 ------cga---attgg---tcaaagcctact----gcaaagccccgaac
SEQ ID NO:20    834 ------ctatagcgtag-aggacattgcaaaa----agata----cttta
SEQ ID NO:21    827 cgaaagcga---ccgagcaatgcgtggcagaaacgggcgatgtctacgac

SEQ ID NO:17    999 atac------atcaactgctgcct----------cgaacagcgcggtgct
SEQ ID NO:19   1159 -tcc------gtca------gcat----------cgaacagggtgttgcc
SEQ ID NO:20    869 aaat------cccatgtgcttgtagatttaaaaacgatgagagagttgaa
SEQ ID NO:21    874 gcgctggcggataaatatctggc------------gattggctgctcct

SEQ ID NO:17   1033 gttcttgaaaaagttgtccgcgatggcaaatgcgacggc-ttgatcatgc
SEQ ID NO:19   1186 tggcgtgaaggcctgatccgcgacaacaaggttgacggc-gtactggttc
SEQ ID NO:20    913 aatataaagagattggttaaagagttggacgtcgatggagttgtttat--
SEQ ID NO:21    911 gtgtttcgccgaacgatcagcgcctgaaaatgc-tcagc-cagatggtgg

SEQ ID NO:17   1082 accagaacc-gttcctgcaagaacatgagcctcctcaacaacgaaggcg-
SEQ ID NO:19   1235 actacaacc-ggtcctgcaaaccctggagcggctacatgcctgaaatgc-
SEQ ID NO:20    961 ----tacac-tttgcagtattgccat----acatttaacatagagggagc
SEQ ID NO:21    959 aggaatatcaggtcgatggcgtagttga-----tgtgattttgcaggcgt

SEQ ID NO:17   1130 ---gccagcgcatc-cagaagaacctc--ggcgtaccgtacgtcatcttc
SEQ ID NO:19   1283 ---agcgtcgtttc-accaaagacatg--ggtatccccactgctggattc
SEQ ID NO:20   1002 taaggtagaggagg-cattaaaagaggagggcattccaattataagaatt
SEQ ID NO:21   1004 ---gccatacctacgcggtggaatcgc--tggcgattaaacgtcatgtgc

SEQ ID NO:17   1174 gacggcgaccagaccgatgctcgtaacttctcggaagca-----------
SEQ ID NO:19   1327 gacggtgaccaggctgacccgagaaacttcaacgcggct-----------
SEQ ID NO:20   1051 gaaactgactattctga-------aagtgatag--agag-----------
SEQ ID NO:21   1049 gccagc-agcacaacattccttatatcgctattgaaacagactactccac

SEQ ID NO:17   1213 ------------cagttcgatacccgcgtagaagctttggcagaaatga
SEQ ID NO:19   1366 ------------cagtatgagacccgtgttcagggcttggtcgaagcca
SEQ ID NO:20   1081 ------------cagttaaaaacaaggttggaggcatttattgagatga
SEQ ID NO:21   1098 ctcggatgtcgggcagctcagtacccgtgtcgcggcctttattgagatgc

SEQ ID NO:17   1250 tggcagacaaaaaagccaatgaaggaggaaaccactaa
SEQ ID NO:19   1403 tggaag-caaatgatgaaaagaagg-ggaaataa----
SEQ ID NO:20   1118 t----------------ttaa--------------
SEQ ID NO:21   1148 tgtaa-------------------------------
```

**Figure 17**

```
SEQ ID NO:18      1 --mseektvdiesmsskealgyflpkvdedarkakkegrlvcwsasvapp
SEQ ID NO:22      1 ----mpktvs----pgvqalrdvvekvyrelrepkergekvgwssskfpc
SEQ ID NO:23      1 --mmklka--ieklmqkfa------srkeqlykqkeegrkvfgm------
SEQ ID NO:24      1 mslvtdlpaifdqfsearqtg-fltvmdlkergiplvg------------

SEQ ID NO:18     49 efctamdiaivypethaag---igarhgapamlevaenkgynqdicsycr
SEQ ID NO:22     43 elaesfrlhvgypenqaag---iaanrdgevmcqaaedigydndicgyar
SEQ ID NO:23     35 -fcayvpieiila-anaip---vglcggkndtipiae-edlprnlcplik
SEQ ID NO:24     38 tyctfmpqei----pmaagavvvslcstsdetieeae-kdlprnlcplik

SEQ ID NO:18     96 vnmgym----------------------------------------------
SEQ ID NO:22     90 islayaagfrgankmdkdgnyvinphsgkqmkdangkkvfdadgkpvidp
SEQ ID NO:23     79 ssygf-----------------------------------------------
SEQ ID NO:24     83 ssygf-----------------------------------------------

SEQ ID NO:18    102 ellkqqaltgetpev---------------lknspaspiplpdvvltcnn
SEQ ID NO:22    140 ktlkpfattdniyeiaalpegeektrrqnalhkyrqmtmpmpdfvlccnn
SEQ ID NO:23     84 -------------------------------kkaktcpyfeasdiviget
SEQ ID NO:24     88 ------gktdkcpyf---------------y---------fsdlvvg-et

SEQ ID NO:18    137 icntllkwyenlakelnvplinidvpfnhefpvtkhakqyivgefkhaik
SEQ ID NO:22    190. icncmtkwyediarrhniplimidvpynefdhvneanvkyirsqldtair
SEQ ID NO:23    103 tcegkkkmfelm--erlvpmhimhlphmkd----edslkiwikeveklke
SEQ ID NO:24    107 tcdgkkkmyeymaefkpvhvmqlpnsvkdd-----asralwkaemlrlqk

SEQ ID NO:18    187 qledlcgrpfdydkffe---vqkqtqrsiaawnkiatyfqykpsplngfd
SEQ ID NO:22    240 qmeeitgkkfdedkfeq---ccqnanrtakawlkvcdylqykpapfngfd
SEQ ID NO:23    147 lveketgnkiteeklke---tvdkvnkvrelfyklyelrknkpapikgld
SEQ ID NO:24    152 tveerfgheisedalrdaialknrerralanfyhlg---qlnppalsgsd

SEQ ID NO:18    234 ---lfnymglavaarslnyseitfnkflkeldekvan--kkwafge--n-
SEQ ID NO:22    287 ---lfnhmadvvtargrveaaeafellakeleqhvke--gtttapf--k-
SEQ ID NO:23    194 vlklfqfaylldiddtigile----dlieeleerv----kk---ge--gy
SEQ ID NO:24    199 ---ilk----vvygatfrfdk---ealineldamtarvrqqweegqrld-

SEQ ID NO:18    276 eksrvtwegiavwialghtfkelkgqgalmtg----say---pgmwdvsy
SEQ ID NO:22    329 eqhrimfegipcwpklpnlfkplkanglnitg----vvy---apafgfvy
SEQ ID NO:23    231 egkrilitgcpmvagnnkiveiieevggvvvg----eesctgtrffenfv
SEQ ID NO:24    238 prprilitgcpiggaaekvvraieenggwvvgyenctga---kateqcva

SEQ ID NO:18    319 epgdl-esmaeaysrtyinccl--eqrgavlekvvrdgkcdglimhqnrs
SEQ ID NO:22    372 --nnl-delvkayckapnsvsi--eqgvawreglirdnkvdgvlvhynrs
SEQ ID NO:23    277 egysv-ediakryfkipcacrfkndervenikrlvkeldvdgvvyytlqy
SEQ ID NO:24    285 etgdvydaladkylaigcscvspndqrlkmlsqmveeyqvdgvvdvilqa

SEQ ID NO:18    366 cknmsllnnegg--qriqknlgvpyvifdgdqtdarnfseaqfdtrveal
SEQ ID NO:22    417 ckpwsgympemq--rrftkdmgiptagfdgdqadprnfnaaqyetrvqgl
SEQ ID NO:23    326 cht---fniegakveealkeegipiirietdyses---dreqlktrleaf
SEQ ID NO:24    335 chtyaveslaik--rhvrqqhnipyiai---etdystsdvgqlstrvaaf

SEQ ID NO:18    414 aemmadkkaneggnh
SEQ ID NO:22    465 veameandekkgk--
SEQ ID NO:23    370 iemi-----------
SEQ ID NO:24    380 ieml-----------
```

**Figure 18**

ATGAGTCAGATCGACGAACTTATCAGCAAATTACAGGAAGTATCCAACCATCCCCAGAAG
ACGGTTTTGAATTATAAAAAACAGGGTAAAGGCCTCGTAGGCATGATGCCCTACTACGCT
CCGGAAGAAATCGTATATGCTGCAGGCTACCTCCCGGTAGGCATGTTCGGTTCCCAGAAC
CCGCAGATCTCCGCAGCTCGTACGTACCTTCCTCCGTTCGCTTGCTCCTTGATGCAGGCT
GACATGGAACTCCAGCTCAACGGCACCTATGACTGCCTCGACGCTGTTATCTTCTCCGTT
CCTTGCGACACTCTCCGCTGCATGAGCCAGAAATGGCACGGCAAAGCTCCGGTCATCGTC
TTCACACAGCCGCAGAACCGTAAGATCCGCCCGGCTGTCGATTTCCTCAAAGCTGAATAC
GAACATGTCCGTACGGAATTGGGACGTATCCTCAACGTAAAAATCTCCGACCTGGCTATC
CAGGAAGCTATCAAAGTATATAACGAAAACCGTCAGGTTATGCGTGAATTCTGCGACGTA
GCTGCTCAGTACCCGCAGATCTTCACTCCGATAAAACGTCATGACGTCATCAAAGCCCGC
TGGTTCATGGACAAAGCTGAACACACCGCTTTGGTCCGCGAACTCATCGACGCTGTCAAG
AAAGAACCGGTACAGCCGTGGAATGGCAAAAAAGTCATCCTCTCCGGTATCATGGCAGAA
CCGGATGAATTCCTCGATATCTTCAGCGAATTCAACATCGCTGTCGTCGCTGACGACCTC
GCTCAGGAATCCCGCCAGTTCCGTACAGACGTACCGTCCGGCATCGATCCCCTCGAACAG
CTCGCTCAGCAGTGGCAGGACTTCGATGGCTGCCCGCTCGCTTTGAACGAAGACAAACCG
CGTGGCCAGATGCTCATCGACATGACTAAGAAATACAATGCTGACGCCGTCGTCATCTGC
ATGATGCGTTTCTGCGATCCTGAAGAATTCGACTATCCGATTTACAAACCGGAATTTGAA
GCTGCTGGCGTTCGTTACACGGTCCTCGACCTCGACATCGAATCTCCGTCCCTCGAACAG
CTCCGCACCCGTATCCAGGCTTTCTCGGAAATCCTCTAA (SEQ ID NO:25)

**Figure 19**

MSQIDELISKLQEVSNHPQKTVLNYKKQGKGLVGMMPYYAPEEIVYAAGYLPVGMFGSQN
PQISAARTYLPPFACSLMQADMELQLNGTYDCLDAVIFSVPCDTLRCMSQKWHGKAPVIV
FTQPQNRKIRPAVDFLKAEYEHVRTELGRILNVKISDLAIQEAIKVYNENRQVMREFCDV
AAQYPQIFTPIKRHDVIKARWFMDKAEHTALVRELIDAVKKEPVQPWNGKKVILSGIMAE
PDEFLDIFSEFNIAVVADDLAQESRQFRTDVPSGIDPLEQLAQQWQDFDGCPLALNEDKP
RGQMLIDMTKKYNADAVVICMMRFCDPEEFDYPIYKPEFEAAGVRYTVLDLDIESPSLEQ
LRTRIQAFSEIL (SEQ ID NO:26)

**Figure 20**

```
SEQ ID NO:25      1 atgagtcagatcgacgaacttatcagcaaattacaggaagtatccaacca
SEQ ID NO:27      1 atggct---atcagtgcacttattgaagagttccaaaaagtat-ctgcca
SEQ ID NO:28      1 ----------atgatgaaattaaaggcaattgaaaagttgatgcaaaaat
SEQ ID NO:29      1 atgtcacttgtcaccgatctacccgccattttcgatcagttctctgaagc

SEQ ID NO:25     51 tccccagaag---------ac-------ggttttg---aattataaaaaa
SEQ ID NO:27     47 gccc--gaag---------ac-------catgctggccaaatataaagcc
SEQ ID NO:28     41 tcgccagtag---------aaaagaacagctatat---aagcaaaaagaa
SEQ ID NO:29     51 tcgccagacaggctttctcac-------cgtcatg---gatctcaaggag

SEQ ID NO:25     82 cagggtaaaggcctcgtaggca--tgatgccctactacgctccggaagaa
SEQ ID NO:27     79 cagggcaaaaaagccatcggct--gcctgccgtactatgttccggaagaa
SEQ ID NO:28     79 gaaggtagaaaagtttttggaa--tgttctgtgcctatgttccaatagaa
SEQ ID NO:29     91 cgcggcattccgctggttggcacttactgcacctttatgc--cgcaagag

SEQ ID NO:25    130 atcgtatatgctgcaggctacctcccggtaggcatgt---tcggttccca
SEQ ID NO:27    127 ctggtctatgctgcaggcatggttcccatgggtgtat---ggggctgcaa
SEQ ID NO:28    127 ataattttagcagcaaatgcaatcccagttggtttgt---gtggaggtaa
SEQ ID NO:29    139 atcccgatggcagccgg------tgcggttgtggtttcgctctgttccac

SEQ ID NO:25    177 -------gaacccgcag-atctccgcagctcgtacgtaccttcctccgtt
SEQ ID NO:27    174 -------tggcaaacaggaagtccgttccaaggaa-tactgtgcttcctt
SEQ ID NO:28    174 -------aaatgacaca-atcccaatagcagaggaggatttgccaagaaa
SEQ ID NO:29    183 ctctgatgaaacc-----attgaagaagcggagaaagatctgccgcgcaa

SEQ ID NO:25    219 cgcttgctccttgatgcaggctgacatggaactccagctcaacggca---
SEQ ID NO:27    216 ctactgcaccattgcccagcagtctctggaaatgctgctggacggga---
SEQ ID NO:28    216 cctatgcccattaataaaatcatcctatggttttaag----aaggca---
SEQ ID NO:29    228 cctctgcccgctga-------ttaaaagcagctacggct--tcggcaaaa

SEQ ID NO:25    266 cctatgactgcctcgacgctgttatcttctcc----gttcct-tgcg---
SEQ ID NO:27    263 ccctggatgggttggacgggatcatca-ctcc----ggtactgtgtg---
SEQ ID NO:28    259 --aaaacctgcccttactttg-aagcatctgatatagttatt-ggag---
SEQ ID NO:29    269 ccgataaatgcccctac----ttctacttttc----ggatct-ggtggtc

SEQ ID NO:25    308 ----acactctccgctgcatgagccagaaat-------gg--------c-
SEQ ID NO:27    305 ----ataccctgcgtcccatgagccagaacttcaaagtgg--------cc
SEQ ID NO:28    302 ----aaact-------------acctgtgaa-------gg--------a-
SEQ ID NO:29    310 ggtgaaaccacctgcgacggcaaaaagaaaa-------tgtatgaatac-

SEQ ID NO:25    338 ----acggcaaagct----ccggtcatcg-tcttcacacagccgcagaac
SEQ ID NO:27    343 atgaaagacaagatg----ccggttattt-tcctggctcatccccaggtc
SEQ ID NO:28    319 ----aagaagaagat----gtttgagttgatggagagattggtgccaatg
SEQ ID NO:29    352 ----atggcggagtttaagcctgttcatg-tgatgcaattgcccaacagc

SEQ ID NO:25    379 cgtaaga-tccgcccggc------------tgtcgatttcctcaaag-ct
SEQ ID NO:27    388 cgtcagaatgccgccggc------------aagc-agttcacctatg-at
SEQ ID NO:28    361 catataa-tgcacctcccacacatgaaagatgaagattctttgaaaatct
SEQ ID NO:29    397 gttaagg-acgatgcctc------------gcgtgcgttatggaaag-cc

SEQ ID NO:25    415 gaat--acgaacatgtc---cgt-------acgg--aattgg----gacg
SEQ ID NO:27    424 gcct--acagcgaagt-----ga-------aaggccatctgg----aaga
SEQ ID NO:28    410 ggattaaagaagttgaaaagcta-------aaag--aattggttgagaaa
SEQ ID NO:29    433 ga------gatgctgcg---cttgcaaaaaacgg--tagaag----aacg
```

```
SEQ ID NO:25    447 tatcctcaacgtaaaa--atctccgacctggctatccaggaagctatcaa
SEQ ID NO:27    456 aatctgcggccatgaa--atcaccaatgatgccatcctggatgccatcaa
SEQ ID NO:28    451 gagactggaaataaaataacagaggaaaagttaaaagagacagttgataa
SEQ ID NO:29    468 ttttgggcacg---ag--attagcgaagatgctctgcgcgatgccattgc

SEQ ID NO:25    495 agtatataacgaaaaccgtcaggttatgcgtgaattct---------gcg
SEQ ID NO:27    504 agtgtacaacaagagccgtgctgcccgccgcgaattct---------gca
SEQ ID NO:28    501 agtaaataaagtta---gggagttgttttataaactct---------atg
SEQ ID NO:29    513 gctgaaaaaccgcgaacgtcgcgcactggctaatttttatcatcttgggc

SEQ ID NO:25    536 acgtagctgctcag-----tacccgcagatcttcactccgataaa--acg
SEQ ID NO:27    545 aactggc--caacg-----aacatcctgatctgatcccggcttccgtacg
SEQ ID NO:28    539 a-attgaggaagaa-----taaac-cag-------ctccaattaa--ggg
SEQ ID NO:29    563 agttaaatcctccggcgcttagcggcag--cgacattctgaaagt--ggt

SEQ ID NO:25    579 tcatgacgtcatc------------aaag----cccgctgg-----ttca
SEQ ID NO:27    588 ggccaccgtactg------------cgtg----ccgcttac-----ttca
SEQ ID NO:28    573 tttagatgtttta------------aaattattccagtttgcctatttat
SEQ ID NO:29    609 ttacggcgcaaccttccggttcgataaag----aggcgttg-----atca

SEQ ID NO:25    608 tggacaaagctgaacacaccgctttggtccgcgaactcatcgacgctgtc
SEQ ID NO:27    617 tgctgaaggatgaatacaccgaaaagctggaagaactgaacaagg-----
SEQ ID NO:28    611 tggatattgatgacacaatagggattttagaggatttaattgaggagtta
SEQ ID NO:29    650 atgaactggatgcaatgaccgc------ccgcg--ttcgtcagcagtggg

SEQ ID NO:25    658 aagaa-------ag--aaccggtacagccgtggaat-----ggcaaaaaa
SEQ ID NO:27    662 aactg-------gc--agctgctcctgccggcaagttcgacggccacaaa
SEQ ID NO:28    661 gaggagagagttaa--aaaaggagaaggttatgaa------ggaaagaga
SEQ ID NO:29    692 aagaa-------ggccagcgactggacccgcgtccg-----cgcatttta

SEQ ID NO:25    694 gtcatcctctccggt------atcatggcagaaccggatgaattcct---
SEQ ID NO:27    703 gtggttgtttccggc------atcatctacaacacgcccggcatcct---
SEQ ID NO:28    703 attttaataactggctgtccaatggttgctggaaacaataagattgt---
SEQ ID NO:29    730 atcaccggctgcccg------attggcggcgcagcagaaaaagtggtgcg

SEQ ID NO:25    735 cgatatcttcagcgaatt-caacatcgctgtcgtcgctgacgacctc-gc
SEQ ID NO:27    744 gaaagccatggatgacaa-caaactggccattgctgctgatgactgc-gc
SEQ ID NO:28    750 tgaaattattgaggaagt-tggaggagtagttgttggtgaagaaagctgc
SEQ ID NO:29    774 cgcgat-tgaagagaatggcggctgggttgtcggttatgaaaactgc-ac

SEQ ID NO:25    783 tcagga-atcccgccagttccgtacagacgtaccgtccggcatcgatccc
SEQ ID NO:27    792 ttatga-aagccgcagctttgccgtggatgctccggaagatctgga---c
SEQ ID NO:28    799 actgga-a-----caagattctttgaaaactttgttgagg--gctatagc
SEQ ID NO:29    822 cggggcgaaagcgaccgagcaatgc-gtggcagaaacggg---cgatgtc

SEQ ID NO:25    832 ctcgaacagctcgctcag------cagtgg------caggacttcgat-g
SEQ ID NO:27    838 aacggactgcatgctctggctgtacagttctccaaacagaagaacgat-g
SEQ ID NO:28    841 gtagaggacattgcaaa--------------------aaga-tacttt-a
SEQ ID NO:29    868 tacgacgcgctggcggat------aaatat------ctgg----cgattg

SEQ ID NO:25    869 ----gctgcccgctcgctttgaa----cgaagacaaaccgcg-tggccag
SEQ ID NO:27    887 ttctgctgtacgatcc---tgaatttgccaagaatacccgttctgaacac
SEQ ID NO:28    869 ----aaatcccatgtgcttgta------gatttaaaaacgat-gagagag
SEQ ID NO:29    902 ----gctgctc-ctgtgtttcgc----cga--acgatcagcg-cctgaaa

SEQ ID NO:25    910 atgctcatcgaca-----tgactaagaaatacaatgctgacgccgtcgtc
SEQ ID NO:27    934 gttggca---atc-----tggtaaaagaaagcggcgcagaaggactgatc
SEQ ID NO:28    908 ttgaaaatataaagagattggttaaagagttggacgtcgatggagttgtt
SEQ ID NO:29    940 atgctcagccaga-----tggtggaggaatatcaggtcgatggcgtagtt
```

```
SEQ ID NO:25    955 atctgcatgatgcgtttctgcgatcctgaagaattcgactatc---cgat
SEQ ID NO:27    976 gtgttcatgatgcagttctgcgatccggaagaaatggaatatc---ctga
SEQ ID NO:28    958 tattacactttgcagtattgccatacatttaacatagagggag---ctaa
SEQ ID NO:29    985 gatgtgattttgcaggcgtgccatacctacgcggtggaatcgctggcgat

SEQ ID NO:25   1002 ttacaaaccggaatttgaagctgctgg----cgttcgttacacggtcctc
SEQ ID NO:27   1023 tctgaagaaggctctggatgcccacca----cattcctcatgtgaagatt
SEQ ID NO:28   1005 ggtagaggaggcattaaaagaggaggg----cattc------caattata
SEQ ID NO:29   1035 t----aaacgtcatgtgcgccagcagcacaacattccttatatcgctatt

SEQ ID NO:25   1048 gacctcgacatcgaatctccgtccctcgaa------cagctccgcacccg
SEQ ID NO:27   1069 ggtgtggaccagatgacccgggactttggt------caggcccagaccgc
SEQ ID NO:28   1045 agaattgaaactgactattctgaaagtgatagagagcagttaaaaacaag
SEQ ID NO:29   1081 gaaacagactactccacctcggatgtcggg------cagctcagtacccg

SEQ ID NO:25   1092 tatccaggctttctcggaaatcctctaa
SEQ ID NO:27   1113 tctggaagctttcgcagaaagcctgtaa
SEQ ID NO:28   1095 gttggaggcatttattgagatgatttaa
SEQ ID NO:29   1125 tgtcgcggcctttattgagatgctgtaa
```

## Figure 21

```
SEQ ID NO:26      1 msqideliskqevsnhpqk---tvlnykkqgkglvgmmpyyapeeivya
SEQ ID NO:30      1 -maisalieefqkvsaspkt---mlakykaqgkkaigclpyyvpeelvya
SEQ ID NO:31      1 mmkl-kaieklmqkfasrke---qlykqkeegrkvfgmfcayvpieiila
SEQ ID NO:32      1 mslvtdlpaifdqfsearqtgfltvmdlkergiplvgtyctfmpqeipma

SEQ ID NO:26     48 agylpvgmfgsqnpqisaartylppfacslmqadmelqlngt---ydc--
SEQ ID NO:30     47 agmvpmgvwgcngkqevrskeycasfyctiaqqslemlldgt---ldg--
SEQ ID NO:31     47 anaipvglcggkndtipiaeedlprnlcplikssygfkkaktcpyfea--
SEQ ID NO:32     51 agavvvslcstsdetieeaekdlprnlcplikss--ygfgkt---dkcpy

SEQ ID NO:26     93 ---ldavifsvpcdtlrcmsqkwh----gkapvivftqpqnrkirpavdf
SEQ ID NO:30     92 ---ldgiitpvlcdtlrpmsqnfkvamkdkmpviflahpqvrqnaagkqf
SEQ ID NO:31     95 ---sdivigettcegkkkmfelme----rlvpmhimhlp-hmkdedslki
SEQ ID NO:32     96 fyfsdlvvgettcdgkkkmyeyma----efkpvhvmqlpnsvkddasral

SEQ ID NO:26    136 lkaeyehvrtelgrilnvkisdlaiqeaikvynenrqvmrefcdvaaqyp
SEQ ID NO:30    139 tydaysevkghleeicgheitndaildaikvynksraarrefcklanehp
SEQ ID NO:31    137 wikeveklkelveketgnkiteeklketvdkvnkvrelfyklyelrknkp
SEQ ID NO:32    142 wkaemlrlqktveerfgheisedalrdaialknrerralanfyhlgqlnp

SEQ ID NO:26    186 qiftpikrhdvik----arwf---mdkaehtalvrelidavkk--epvqp
SEQ ID NO:30    189 dlipasvratvlr----aayf---mlkdeytekleelnkelaa--apagk
SEQ ID NO:31    187 ---apikgldvlk----lfqfaylldiddtigiledlieeleervkkgeg
SEQ ID NO:32    192 ---palsgsdilkvvygatfr---fdk---ealinel-damta--rvrqq

SEQ ID NO:26    227 wn-gkk------vilsg--imaepdefldifsefniavvaddlaqesrqf
SEQ ID NO:30    230 fd-ghk------vvvsg--iiyntpgilkamddnklaiaaddcayesrsf
SEQ ID NO:31    230 ye-gkr------ilitgcpmvagnnkiveiieevggvvvgeesctgtrff
SEQ ID NO:32    230 weegqrldprprilitgcpiggaaekvvraieenggwvvgyenctgakat

SEQ ID NO:26    268 rtdvpsgidp-leqlaqqwqdfdgcplalned---kprgqmlidmtkkyn
SEQ ID NO:30    271 avdapedldnglhalavqfskqkndvllydpefakntrsehvgnlvkesg
SEQ ID NO:31    273 enfv-egys--vediakryfkip-cacrfknd---e-rvenikrlvkeld
SEQ ID NO:32    280 eqcvaetgdv-ydaladkylai-gcscvspnd---q-rlkmlsqmveeyq

SEQ ID NO:26    314 adavvicmmrfcdpeefdypiykpef-eaagvrytvldldiespsleqlr
SEQ ID NO:30    321 aeglivfmmqfcdpeemeypdlkkal-dahhiphvkigvdqmtrdfgqaq
SEQ ID NO:31    315 vdgvvyytlqychtfniegakveeal-keegipiirietdysesdreqlk
SEQ ID NO:32    324 vdgvvdvilqachtyaveslaikrhvrqqhnipyiaietdystsdvgqls

SEQ ID NO:26    363 triqafseil
SEQ ID NO:30    370 taleafaesl
SEQ ID NO:31    364 trleafiemi
SEQ ID NO:32    374 trvaafieml
```

## Figure 22

```
   1  CGACGGCCCG GGCTGGTATC ATTCTAGTCA GTAATTCACC TTTGGAAAAT TTTCACAAAG
  61  GCAGTACGAC AGAAGCGTCG ATACATTCCA TTTAGCAGGA GGAAGTTACG GTAATGAGAA
 121  AAGTAGAAAT CATTACAGCT GAACAAGCAG CTCAGCTCGT AAAAGACAAC GACACGATTA
 181  CGTCTATCGG CTTTGTCAGC AGCGCCCATC CGGAAGCACT GACCAAAGCT TTGGAAAAAC
 241  GGTTCCTGGA CACGAACACC CCGCAGAACT TGACCTACAT CTATGCAGGC TCTCAGGGCA
 301  AACGCGATGG CCGTGCCGCT GAACATCTGG CACACACAGG CCTTTTGAAA CGCGCCATCA
 361  TCGGTCACTG GCAGACTGTA CCGGCTATCG GTAAACTGGC TGTCGAAAAC AAGATTGAAG
 421  CTTACAACTT CTCGCAGGGC ACGTTGGTCC ACTGGTTCCG CGCCTTGGCA GGTCATAAGC
 481  TCGGCGTCTT CACCGACATC GGTCTGGAAA CTTTCCTCGA TCCCCGTCAG CTCGGCGGCA
 541  AGCTCAATGA CGTAACCAAA GAAGACCTCG TCAAACTGAT CGAAGTCGAT GGTCATGAAC
 601  AGCTTTTCTA CCCGACCTTC CCGGTCAACG TAGCTTTCCT CCGCGGTACG TATGCTGATG
 661  AATCCGGCAA TATCACCATG GACGAAGAAA TCGGGCCTTT CGAAAGCACT TCCGTAGCCC
 721  AGGCCGTTCA CAACTGTGGC GGTAAAGTCG TCGTCCAGGT CAAAGACGTC GTCGCTCACG
 781  GCAGCCTCGA CCCGCGCATG GTCAAGATCC CTGGCATCTA TGTCGACTAC GTCGTCGTAG
 841  CAGCTCCGGA AGACCATCAG CAGACGTATG ACTGCGAATA CGATCCGTCC CTCAGCGGTG
 901  AACATCGTGC TCCTGAAGGC GCTACCGATG CAGCTCTCCC CATGAGCGCT AAGAAAATCA
 961  TCGGCCGCCG CGGCGCTTTG GAATTGACTG AAAACGCTGT CGTCAACCTC GGCGTCGGTG
1021  CTCCGGAATA CGTTGCTTCT GTTGCCGGTG AAGAAGGTAT CGCCGATACC ATTACCCTGA
1081  CCGTCGAAGG TGGCGCCATC GGTGGCGTAC CGCAGGGCGG TGCCCGCTTC GGTTCGTCCC
1141  GCAATGCCGA TGCCATCATC GACCACACCT ATCAGTTCGA CTTCTACGAT GGCGGCGGTC
1201  TGGACATCGC TTACCTCGGC CTGGCCCAGT GCGATGGCTC GGGCAACATC AACGTCAGCA
1261  AGTTCGGTAC TAACGTTGCC GGCTGCGGCG GTTTCCCCAA CATTTCCCAG CAGACACCGA
1321  ATGTTTACTT CTGCGGCACC TTCACGGCTG GCGGCTTGAA AATCGCTGTC GAAGACGGCA
1381  AAGTCAAGAT CCTCCAGGAA GGCAAAGCCA AGAAGTTCAT CAAAGCTGTC GACCAGATCA
1441  CTTTCAACGG TTCCTATGCA GCCCGCAACG GCAAACACGT TCTCTACATC ACAGAACGCT
1501  GCGTATTTGA ACTGACCAAA GAAGGCTTGA AACTCATCGA AGTCGCACCG GGCATCGATA
1561  TTGAAAAAGA TATCCTCGCT CACATGGACT TCAAGCCGAT CATTGATAAT CCGAAACTCA
1621  TGGATGCCCG CCTCTTCCAG GACGGTCCCA TGGGACTGAA AAAATAAATC TCTGCTGTAA
1681  AGGAGACTTT ACTATGAAAC CAATGAGACT ACATCACGTA GGCATTGTCC TGCCGACCTT
1741  AGAAAAAGCC CATGAATTCA TGCAGAATAA TGGACTTGAA ATCGACTATG CCGGCTATGT
1801  CGATGCTTAC CAGGCTGATC TCATTTTCAC TAAGTTTGGT GAATTTGCCA GCCCGATTGA
1861  AATGATTATC CCGCACTCCG GTGTGCTTAC CCAATTCAAT GGTGGCCGCG GCGGCATTGC
1921  CCACATCGCC TTCGAAGTGG ACGATGTCGA AGCTGTCCGC CAGGAAATGG AAGCAGATTG
1981  TCCGGGATGC ATGTTAGAAA AGAAAGCTGT CCAGGGTACG GACGACATTA TCGTCAACTT
2041  CCGCCGCCCG ACAACCAACC AGGGTATCCT CGTTGAATAT GTTCAGACGA CAGCACCTAT
2101  CACCGGCCGC GGCGAAAATC CTTTCGTTAA GAATCTCGGC CCGGAAAAAG GGAAGCTCAA
2161  CGAAACATGG CATCCCATGC GCCTGCACCA TATCGGCATC GTCTTGCCGA CCTTGGAAAA
2221  GGCCCATGAA TTCATCAAGA CCAATGGTCT GGAAGTGGAT TATTCCGGTT TCGTCGACGC
2281  CTACCATGCG GATCTCATTT TCACTAAAAA AGGTGAAAAC AGTACGCCTA TCGAATTCAT
2341  TATTCCCCGT GAAGGGGTCC TCAAAGATTT CAATCATGGC AGGGGAGGTA TCGCTCATAT
2401  CGCCTTTGAA GTGGATGATG TCGAAAAGGT ACGTCAGATT ATGGAAAGCC AGAAGCCTGG
2461  TTGCATGCTC GAAAAGAAAG CCGTCCGGGG AACGGACGAT ATCATCGTCA ACTTCCGCCG
2521  TCCCAGCACG GACGCCGGCA TCCTCGTCGA ATATGTCCAG ACCGTAGCTC CCATCAATCG
2581  CAGCAATCCC AACCCTTTTA ATGATTGATT TTTTATAAAG AAAGGTGAAA ACTGTGTATA
2641  CTCTCGGAAT CGACGTTGGT TCTTCTTCTT CCAAGGCAGT CATCCTGGAA GATGGCAAGA
2701  AGATCGTCGC CCATGCCGTC GTTGAAATCG GCACCGGTTC GACCGGTCCG GAACGCGTCC
2761  TGGACGAAGT CTTCAAAGAT ACCAACTTAA AAATTGAAGA CATGGCGAAC ATCATCGCCA
2821  CAGGCTATGG CCGTTTCAAT GTCGACTGCG CCAAAGGCGA AGTCAGCGAA ATCACGTGCC
2881  ATGCCAAAGG GGCCCTCTTT GAATGCCCCG GTACGACGAC CATCCTCGAT ATCGGCGGTC
2941  AGGACGTCAA GTCCATCAAA TTGAATGGCC AGGGCCTGGT CATGCAGTTT GCCATGAACG
3001  ACAAATGCGC CGCTGGTACG GGCCGTTTCC TCGACGTCAT GTCGAAGGTA CTGGAAATCC
3061  CCATGTCTGA AATGGGGGAC TGGTACTTCA AATCGAAGCA TCCCGCTGCC GTCAGCAGTA
3121  CCTGCACGGT TTTTGCTGAA TCGGAAGTCA TTTCCCTTCT TTCCAAGAAT GTCCCGAAAG
3181  AAGATATCGT AGCCGGTGTC CATCAGTCCA TCGCCGCCAA AGCCTGCGCT CTCGTGCGCC
3241  GCGTCGGTGT CGGTGAAGAC CTGACCATGA CCGGCGGTGG CTCCCGCGAT CCCGGCGTCG
3301  TCGATGCCGT ATCGAAAGAA TTAGGTATTC CTGTCAGAGT CGCTCTGCAT CCCCAAGCGG
3361  TGGGTGCTCT CGGAGCTGCT TTGATTGCTT ATGATAAAAT CAAGAAATAA GTCAAAGGAG
```

3421 AGAACAAAAT CATGAGTGAA GAAAAAACAG TAGATATTGA AAGCATGAGC TCCAAGGAAG
3481 CCCTTGGTTA CTTCTTGCCG AAAGTCGATG AAGACGCACG TAAAGCGAAA AAAGAAGGCC
3541 GCCTCGTTTG CTGGTCCGCT TCTGTCGCTC CTCCGGAATT CTGCACGGCT ATGGACATCG
3601 CCATCGTCTA TCCGGAAACT CACGCAGCTG GTATCGGTGC CCGTCACGGT GCTCCGGCCA
3661 TGCTCGAAGT TGCTGAAAAC AAAGGTTACA ACCAGGACAT CTGTTCCTAC TGCCGCGTCA
3721 ACATGGGCTA CATGGAACTC CTCAAACAGC AGGCTCTGAC AGGCGAAACG CCGGAAGTCC
3781 TCAAAAACTC CCCGGCTTCT CCGATTCCCC TTCCGGATGT TGTCCTCACT TGCAACAACA
3841 TCTGCAATAC CTTGCTCAAA TGGTATGAAA ACTTGGCTAA AGAATTGAAC GTACCTCTCA
3901 TCAACATCGA CGTACCGTTC AACCATGAAT TCCCTGTTAC GAAACACGCT AAACAGTACA
3961 TCGTCGGCGA ATTCAAACAT GCTATCAAAC AGCTCGAAGA CCTTTGCGGC CGTCCCTTCG
4021 ACTATGACAA ATTCTTCGAA GTACAGAAAC AGACACAGCG CTCCATCGCT GCCTGGAACA
4081 AAATCGCTAC GTACTTCCAG TACAAACCGT CGCCGCTCAA CGGCTTCGAC CTCTTCAACT
4141 ACATGGGCCT CGCCGTTGCT GCCCGCTCCT TGAACTACTC GGAAATCACG TTCAACAAAT
4201 TCCTCAAAGA ATTGGACGAA AAAGTAGCTA ATAAGAAATG GGCTTTCGGT GAAAACGAAA
4261 AATCCCGTGT TACTTGGGAA GGTATCGCTG TCTGGATCGC TCTCGGCCAC ACCTTCAAAG
4321 AACTCAAAGG TCAGGGCGCT CTCATGACTG GTTCCGCTTA TCCTGGCATG TGGGACGTTT
4381 CCTACGAACC GGGCGACCTC GAATCCATGG CAGAAGCTTA TTCCCGTACA TACATCAACT
4441 GCTGCCTCGA ACAGCGCGGT GCTGTTCTTG AAAAAGTTGT CCGCGATGGC AAATGCGACG
4501 GCTTGATCAT GCACCAGAAC CGTTCCTGCA AGAACATGAG CCTCCTCAAC AACGAAGGCG
4561 GCCAGCGCAT CCAGAAGAAC CTCGGCGTAC CGTACGTCAT CTTCGACGGC GACCAGACCG
4621 ATGCTCGTAA CTTCTCGGAA GCACAGTTCG ATACCCGCGT AGAAGCTTTG GCAGAAATGA
4681 TGGCAGACAA AAAAGCCAAT GAAGGAGGAA ACCACTAATG AGTCAGATCG ACGAACTTAT
4741 CAGCAAATTA CAGGAAGTAT CCAACCATCC CCAGAAGACG GTTTTGAATT ATAAAAAACA
4801 GGGTAAAGGC CTCGTAGGCA TGATGCCCTA CTACGCTCCG GAAGAAATCG TATATGCTGC
4861 AGGCTACCTC CCGGTAGGCA TGTTCGGTTC CCAGAACCCG CAGATCTCCG CAGCTCGTAC
4921 GTACCTTCCT CCGTTCGCTT GCTCCTTGAT GCAGGCTGAC ATGGAACTCC AGCTCAACGG
4981 CACCTATGAC TGCCTCGACG CTGTTATCTT CTCCGTTCCT TGCGACACTC TCCGCTGCAT
5041 GAGCCAGAAA TGGCACGGCA AAGCTCCGGT CATCGTCTTC ACACAGCCGC AGAACCGTAA
5101 GATCCGCCCG GCTGTCGATT TCCTCAAAGC TGAATACGAA CATGTCCGTA CGGAATTGGG
5161 ACGTATCCTC AACGTAAAAA TCTCCGACCT GGCTATCCAG GAAGCTATCA AAGTATATAA
5221 CGAAAACCGT CAGGTTATGC GTGAATTCTG CGACGTAGCT GCTCAGTACC CGCAGATCTT
5281 CACTCCGATA AAACGTCATG ACGTCATCAA AGCCCGCTGG TTCATGGACA AAGCTGAACA
5341 CACCGCTTTG GTCCGCGAAC TCATCGACGC TGTCAAGAAA GAACCGGTAC AGCCGTGGAA
5401 TGGCAAAAAA GTCATCCTCT CCGGTATCAT GGCAGAACCG GATGAATTCC TCGATATCTT
5461 CAGCGAATTC AACATCGCTG TCGTCGCTGA CGACCTCGCT CAGGAATCCC GCCAGTTCCG
5521 TACAGACGTA CCGTCCGGCA TCGATCCCCT CGAACAGCTC GCTCAGCAGT GGCAGGACTT
5581 CGATGGCTGC CCGCTCGCTT TGAACGAAGA CAAACCGCGT GGCCAGATGC TCATCGACAT
5641 GACTAAGAAA TACAATGCTG ACGCCGTCGT CATCTGCATG ATGCGTTTCT GCGATCCTGA
5701 AGAATTCGAC TATCCGATTT ACAAACCGGA ATTTGAAGCT GCTGGCGTTC GTTACACGGT
5761 CCTCGACCTC GACATCGAAT CTCCGTCCCT CGAACAGCTC CGCACCCGTA TCCAGGCTTT
5821 CTCGGAAATC CTCTAAGAAT CGCCTGAATC ATCAAACATC TGGGCGGGAC TCCGAAAGGT
5881 GCCTGCTACA TGATACATTG CCTGTTTTCA GGCAGACAGA TTTGCAGCTT GCGGCCCCCA
5941 TTGTACGGGC TGCAAGCTGT CAATGATGCT TTAAAGACGG CTCTGCCGTT TTTAAATAAA
6001 AACATAAAAC CATATATAAT CTATTAGGAG GAAACTCAAT CATGGAATTC AAACTTTCTG
6061 AATTACAGCA AGATATCGCA AATCTCGCAA AAGATTTCGC AGAAAAAAAA TTAGCTCCCA
6121 CTGTCAAAGA GCGTGACGAA AAAGAAGTTT TCGATCGTGC TATCCTTGAC GAAGTGGGTA
6181 CTCTCGGCCT TCTCGGTATT CCCTGGGAAG AAGAAAACGG CGGCGTAGGC GCTGACTTCC
6241 TCAGCCTCGC AGTTGCTTGC GAAGAAGTAG CTAAAGTTAC CAGCCCGGGC CGTCG (SEQ
ID NO:33)

**Figure 23**

ATGAAACCAATGAGACTACATCACGTAGGCATTGTCCTGCCGACCTTAGAAAAAGCCCAT
GAATTCATGCAGAATAATGGACTTGAAATCGACTATGCCGGCTATGTCGATGCTTACCAG
GCTGATCTCATTTTCACTAAGTTTGGTGAATTTGCCAGCCCGATTGAAATGATTATCCCG
CACTCCGGTGTGCTTACCCAATTCAATGGTGGCCGCGGCGGCATTGCCCACATCGCCTTC
GAAGTGGACGATGTCGAAGCTGTCCGCCAGGAAATGGAAGCAGATTGTCCGGGATGCATG
TTAGAAAAGAAAGCTGTCCAGGGTACGGACGACATTATCGTCAACTTCCGCCGCCCGACA
ACCAACCAGGGTATCCTCGTTGAATATGTTCAGACGACAGCACCTATCACCGGCCGCGGC
GAAAATCCTTTCGTTAAGAATCTCGGCCCGGAAAAAGGGAAGCTCAACGAAACATGGCAT
CCCATGCGCCTGCACCATATCGGCATCGTCTTGCCGACCTTGGAAAAGGCCCATGAATTC
ATCAAGACCAATGGTCTGGAAGTGGATTATTCCGGTTTCGTCGACGCCTACCATGCGGAT
CTCATTTTCACTAAAAAAGGTGAAAACAGTACGCCTATCGAATTCATTATTCCCCGTGAA
GGGGTCCTCAAAGATTTCAATCATGGCAGGGGAGGTATCGCTCATATCGCCTTTGAAGTG
GATGATGTCGAAAAGGTACGTCAGATTATGGAAAGCCAGAAGCCTGGTTGCATGCTCGAA
AAGAAAGCCGTCCGGGGAACGGACGATATCATCGTCAACTTCCGCCGTCCCAGCACGGAC
GCCGGCATCCTCGTCGAATATGTCCAGACCGTAGCTCCCATCAATCGCAGCAATCCCAAC
CCTTTTAATGATTGA (SEQ ID NO:34)

**Figure 24**

MKPMRLHHVGIVLPTLEKAHEFMQNNGLEIDYAGYVDAYQADLIFTKFGEFASPIEMIIP
HSGVLTQFNGGRGGIAHIAFEVDDVEAVRQEMEADCPGCMLEKKAVQGTDDIIVNFRRPT
TNQGILVEYVQTTAPITGRGENPFVKNLGPEKGKLNETWHPMRLHHIGIVLPTLEKAHEF
IKTNGLEVDYSGFVDAYHADLIFTKKGENSTPIEFIIPREGVLKDFNHGRGGIAHIAFEV
DDVEKVRQIMESQKPGCMLEKKAVRGTDDIIVNFRRPSTDAGILVEYVQTVAPINRSNPN
PFND (SEQ ID NO:35)

**Figure 25**

ATGGAATTCAAACTTTCTGAATTACAGCAAGATATCGCAAATCTCGCAAAAGATTTCGCA
GAAAAAAAATTAGCTCCCACTGTCAAAGAGCGTGACGAAAAAGAAGTTTTCGATCGTGCT
ATCCTTGACGAAGTGGGTACTCTCGGCCTTCTCGGTATTCCCTGGGAAGAAGAAAACGGC
GGCGTAGGCGCTGACTTCCTCAGCCTCGCAGTTGCTTGCGAAGAAGTAGCTAAAGTTACC
AGCCCGGGCCGTCG (SEQ ID NO: 36)

**Figure 26**

MEFKLSELQQDIANLAKDFAEKKLAPTVKERDEKEVFDRAILDEVGTLGLLGIPWEEENG
GVGADFLSLAVACEEVAKVTSPGR (SEQ ID NO:37)

**Figure 27**

```
   1  GTGAGCACAC ACTTGATAGC TGATGCCGTC AATGATCAGT TGTTCGTCTA TAGCAGGCTG
  61  AAAGGACATG GGTTTGGTCA CAGTCTGAGC AGTTGCAGGC AGTCAAACAC GTTCGTAACT
 121  ACGCTGTAGA TGATATAAGC AGTATACCAT CTTGCTACGC TCTCGTTGAT CAGGTTGAAT
 181  GCTTTGAGGA AGGTCAGGCG AATAGCCATG CCTCTTGTTT CCAGAACATG GCATGGGGAT
 241  GGATCGACGG TACCCTGTCG GATGCATGCT ATGCGTGGCA TTCATATCAT CAACCAGAAT
 301  TTGATCTTGA ACTACACAGC AATTCTGCGC GTTATGCAAG TGTCTTCGGT CAGATGGTGA
 361  ACAATTCTCA ATTGTTGAGG TCTTGACGAA TTGCGTTATA CACTGTAGGC TATAGTATGC
 421  ACCCCTTGTT ATCTATATCA CAACCGGTCT ATTAGCATTT GCGTCAAGGA GGATGGTCGA
 481  TGATCGACAC TGCGCCCCTT GCCCCACCAC GGGCGCCCCG CTCTAATCCG ATTCGGGATC
 541  GAGTTGATTG GGAAGCTCAG CGCGCTGCTG CGCTGGCAGA TCCCGGTGCC TTTCATGGCG
 601  CGATTGCCCG GACAGTTATC CACTGGTACG ACCCACAACA CCATTGCTGG ATTCGCTTCA
 661  ACGAGTCTAG TCAGCGTTGG GAAGGGCTGG ATGCCGCTAC CGGTGCCCCT GTAACGGTAG
 721  ACTATCCCGC CGATTATCAG CCCTGGCAAC AGGCGTTTGA TGATAGTGAA GCGCCGTTTT
 781  ACCGCTGGTT TAGTGGTGGG TTGACAAATG CCTGCTTTAA TGAAGTAGAC CGGCATGTCA
 841  TGATGGGCTA TGGCGACGAG GTGGCCTACT ACTTTGAAGG TGACCGCTGG GATAACTCGC
 901  TCAACAATGG TCGTGGTGGT CCGGTTGTCC AGGAGACAAT CACGCGGCGG CGCCTGTTGG
 961  TGGAGGTGGT GAAGGCTGCG CAGGTGTTGC GTGATCTGGG CCTGAAGAAG GGTGATCGGA
1021  TTGCTCTGAA TATGCCGAAT ATTATGCCGC AGATTTATTA TACGGAAGCG GCAAAACGAC
1081  TGGGTATTCT GTACACGCCG GTCTTCGGTG GCTTCTCGGA CAAGACTCTT TCCGACCGTA
1141  TTCACAATGC CGGTGCACGA GTGGTGATTA CCTCTGATGG TGCGTACCGC AACGCGCAGG
1201  TGGTGCCCTA CAAAGAAGCG TATACCGATC AGGCGCTCGA TAAGTATATT CCGGTTGAGA
1261  CGGCGCAGGC GATTGTTGCG CAGACCCTGG CCACCTTGCC CCTGACTGAG TCGCAGCGCC
1321  AGACGATCAT CACCGAAGTG GAGGCCGCAC TGGCCGGTGA GATTACGGTT GAGCGCTCGG
1381  ACGTGATGCG TGGGGTTGGT TCTGCCCTCG CAAAGCTCCG CGATCTTGAT GCAAGCGTGC
1441  AGGCAAAGGT GCGTACAGTA CTGGCGCAGG CGCTGGTCGA GTCGCCGCCG CGGGTTGAAG
1501  CTGTGGTGGT TGTGCGTCAT ACCGGTCAGG AGATTTTGTG GAACGAGGGG CGAGATCGCT
1561  GGAGTCACGA CTTGCTGGAT GCTGCGCTGG CGAAGATTCT GGCCAATGCG CGTGCTGCCG
1621  GCTTTGATGT GCACAGTGAG AATGATCTGC TCAATCTCCC CGATGACCAG CTTATCCGTG
1681  CGCTCTACGC CAGTATTCCC TGTGAACCGG TTGATGCTGA ATATCCGATG TTTATCATTT
1741  ACACATCGGG TAGCACCGGT AAGCCCAAGG GTGTGATCCA CGTTCACGGC GGTTATGTCG
1801  CCGGTGTGGT GCACACCTTG CGGGTCAGTT TTGACGCCGA GCCGGGTGAT ACGATATATG
1861  TGATCGCCGA TCCGGGCTGG ATCACCGGTC AGAGCTATAT GCTCACAGCC ACAATGGCCG
1921  GTCGGCTGAC CGGGGTGATT GCCGAGGGAT CACCGCTCTT CCCCTCAGCC GGGCGTTATG
1981  CCAGCATCAT CGAGCGCTAT GGGGTGCAGA TCTTTAAGGC GGGTGTGACC TTCCTCAAGA
2041  CAGTGATGTC CAATCCGCAG AATGTTGAAG ATGTGCGACT CTATGATATG CACTCGCTGC
2101  GGGTTGCAAC CTTCTGCGCC GAGCCGGTCA GTCCGGCGGT GCAGCAGTTT GGTATGCAGA
2161  TCATGACCCC GCAGTATATC AATTCGTACT GGGCGACCGA GCACGGTGGA ATTGTCTGGA
2221  CGCATTTCTA CGGTAATCAG GACTTCCCGC TTCGTCCCGA TGCCCATACC TATCCCTTGC
2281  CCTGGGTGAT GGGTGATGTC TGGGTGGCCG AAACTGATGA GAGCGGGACG ACGCGCTATC
2341  GGGTCGCTGA TTTCGATGAG AAGGGCGAGA TTGTGATTAC CGCCCCGTAT CCCTACCTGA
2401  CCCGCACACT CTGGGGTGAT GTGCCCGGTT TCGAGGCGTA CCTGCGCGGT GAGATTCCGC
2461  TGCGGGCCTG GAAGGGTGAT GCCGAGCGTT TCGTCAAGAC CTACTGGCGA CGTGGGCCAA
2521  ACGGTGAATG GGGCTATATC CAGGGTGATT TTGCCATCAA GTACCCCGAT GGTAGCTTCA
2581  CGCTCCACGG ACGCCCTGAC GATGTGATCA ATGTGTCGGG CCACCGTATG GGCACCGAGG
2641  AGATTGAGGG TGCCATTTTG CGTGACCGCC AGATCACGCC CGACTCGCCC GTCGGTAATT
2701  GTATTGTGGT CGGTGCGCCG CACCGTGAGA AGGGTCTGAC CCCGGTTGCC TTCATTCAAC
2761  CTGCGCCTGG CCGTCATCTG ACCGGCGCCG ACCGGCGCCG TCTCGATGAG CTGGTGCGTA
2821  CCGAGAAGGG GGCGGTCAGT GTCCCAGAGG ATTACATCGA GGTCAGTGCC TTTCCCGAAA
2881  CCCGCAGCGG GAAGTATATG CGGCGCTTTT TGCGCAATAT GATGCTCGAT GAACCACTGG
2941  GTGATACGAC GACGTTGCGC AATCCTGAAG TGCTCGAAGA GATTGCAGCC AAGATCGCTG
3001  AGTGGAAACG CCGTCAGCGT ATGGCCGAAG AGCAGCAGAT CATCGAACGC TATCGCTACT
3061  TCCGGATCGA GTATCACCCA CCAACGGCCA GTGCGGGTAA ACTCGCGGTA GTGACGGTGA
3121  CAAATCCGCC GGTGAACGCA CTGAATGAGC GTGCGCTCGA TGAGTTGAAC ACAATTGTTG
3181  ACCACCTGGC CCGTCGTCAG GATGTTGCCG CAATTGTCTT CACCGGACAG GGCGCCAGGA
3241  GTTTTGTCGC CGGCGCTGAT ATTCGCCAGT TGCTCGAAGA GATTCATACG GTTGAAGAGG
3301  CAATGGCCCT GCCGAATAAC GCCCATCTTG CTTTCCGCAA GATTGAGCGT ATGAATAAGC
3361  CGTGTATCGC GGCGATCAAC GGTGTGGCGC TCGGTGGTGG TCTGGAATTC GCCATGGCCT
```

```
3421  GCCATTACCG GGTTGCCGAT GTCTATGCCG AATTCGGTCA GCCAGAGATT AATCTGCGCT
3481  TGCTACCTGG TTATGGTGGC ACGCAGCGCT TGCCGCGCCT GTTGTACAAG CGCAACAACG
3541  GCACCGGTCT GCTCCGAGCG CTGGAGATGA TTCTGGGTGG GCGTAGCGTA CCGGCTGATG
3601  AGGCGCTGAA GCTGGGTCTG ATCGATGCCA TTGCTACCGG CGATCAGGAC TCACTGTCGC
3661  TGGCATGCGC GTTAGCCCGT GCCGCAATCG GCGCCGATGG TCAGTTGATC GAGTCGGCTG
3721  CGGTGACCCA GGCTTTCCGC CATCGCCACG AGCAGCTTGA CGAGTGGCGC AAACCAGACC
3781  CGCGCTTTGC CGATGACGAA CTGCGCTCGA TTATCGCCCA TCCACGTATC GAGCGGATTA
3841  TCCGGCAGGC CCATACCGTT GGGCGCGATG CGGCAGTGCA TCGGGCACTG GATGCAATCC
3901  GCTATGGCAT TATCCACGGC TTCGAGGCCG GTCTGGAGCA CGAGGCGAAG CTCTTTGCCG
3961  AGGCAGTGGT TGACCCGAAC GGTGGCAAGC GTGGTATTCG CGAGTTCCTC GACCGCCAGA
4021  GTGCGCCGTT GCCAACCCGC CGACCATTGA TTACACCTGA ACAGGAGCAA CTCTTGCGCG
4081  ATCAGAAAGA ACTGTTGCCG GTTGGTTCAC CCTTCTTCCC CGGTGTTGAC CGGATTCCGA
4141  AGTGGCAGTA CGCGCAGGCG GTTATTCGTG ATCCGGACAC CGGTGCGGCG GCTCACGGCG
4201  ATCCCATCGT GGCTGAAAAG CAGATTATTG TGCCGGTGGA ACGCCCCCGC GCCAATCAGG
4261  CGCTGATCTA TGTTCTGGCC TCGGAGGTGA ACTTCAACGA TATCTGGGCG ATTACCGGTA
4321  TTCCGGTGTC ACGGTTTGAT GAGCACGACC GCGACTGGCA CGTTACCGGT TCAGGTGGCA
4381  TCGGCCTGAT CGTTGCGCTG GGTGAAGAGG CGCGACGCGA AGGCCGGCTG AAGGTGGGTG
4441  ATCTGGTGGC GATCTACTCC GGGCAGTCGG ATCTGCTCTC ACCGCTGATG GGCCTTGATC
4501  CGATGGCCGC CGATTTCGTC ATCCAGGGGA ACGACACGCC AGATGGATCG CATCAGCAAT
4561  TTATGCTGGC CCAGGCCCCG CAGTGTCTGC CCATCCCAAC CGATATGTCT ATCGAGGCAG
4621  CCGGCAGCTA CATCCTCAAT CTCGGTACGA TCTATCGCGC CCTCTTTACG ACGTTGCAAA
4681  TCAAGGCCGG ACGCACCATC TTTATCGAGG GTGCGGCGAC CGGTACCGGT CTGGACGCAG
4741  CGCGCTCGGC GGCCCGGAAT GGTCTGCGCG TAATTGGAAT GGTCAGTTCG TCGTCACGTG
4801  CGTCTACGCT GCTGGCTGCG GGTGCCCACG GTGCGATTAA CCGTAAAGAC CCGGAGGTTG
4861  CCGATTGTTT CACGCGCGTG CCCGAAGATC CATCAGCCTG GGCAGCCTGG GAAGCCGCCG
4921  GTCAGCCGTT GCTGGCGATG TTCCGGGCGC AGAACGACGG GCGACTGGCC GATTATGTGG
4981  TCTCGCACGC GGGCGAGACG GCCTTCCCGC GCAGTTTCCA GCTTCTCGGC GAGCCACGCG
5041  ATGGTCACAT TCCGACGCTC ACATTCTACG GTGCCACCAG TGGCTACCAC TTCACCTTCC
5101  TGGGTAAGCC AGGGTCAGCT TCGCCGACCG AGATGCTGCG GCGGGCCAAT CTCCGCGCCG
5161  GTGAGGCGGT GTTGATCTAC TACGGGGTTG GGAGCGATGA CCTGGTAGAT ACCGGCGGTC
5221  TGGAGGCTAT CGAGGCGGCG CGGCAAATGG GAGCGCGGAT CGTCGTCGTT ACCGTCAGCG
5281  ATGCGCAACG CGAGTTTGTC CTCTCGTTGG GCTTCGGGGC TGCCCTACGT GGTGTCGTCA
5341  GCCTGGCGGA ACTCAAACGG CGCTTCGGCG ATGAGTTTGA GTGGCCGCGC ACGATGCCGC
5401  CGTTGCCGAA CGCCCGCCAG GACCCGCAGG GTCTGAAAGA GGCTGTCCGC CGCTTCAACG
5461  ATCTGGTCTT CAAGCCGCTA GGAAGCGCGG TCGGTGTCTT CTTGCGGAGT GCCGACAATC
5521  CGCGTGGCTA CCCCGATCTG ATCATCGAGC GGGCTGCCCA CGATGCACTG GCGGTGAGCG
5581  CGATGCTGAT CAAGCCCTTC ACCGGACGGA TTGTCTACTT CGAGGACATT GGTGGGCGGC
5641  GTTACTCCTT CTTCGCACCG CAAATCTGGG TGCGCCAGCG CCGCATCTAC ATGCCGACGG
5701  CACAGATCTT TGGTACGCAC CTCTCAAATG CGTATGAAAT TCTGCGTCTG AATGATGAGA
5761  TCAGCGCCGG TCTGCTGACG ATTACCGAGC CGGCAGTGGT GCCGTGGGAT GAACTACCCG
5821  AAGCACATCA GGCGATGTGG GAAAATCGCC ACACGGCGGC CACTTATGTG GTGAATCATG
5881  CCTTACCACG TCTCGGCCTA AAGAACAGGG ACGAGCTGTA CGAGGCGTGG ACGGCCGGCG
5941  AGCGGTAGCG CGGATGGGTA TTGAACAGGT AACGGACGGA AGATCGAACC TTCCGTCCGT
6001  TATCTTTTGG CCGTCGAAGC GTGCTGAGCC GATTATCGTT GCCGTGGTTG TCCCGATGGG
6061  CAGACGCGCT CGAACCAGAT GATACCACCG ACGGCTATCG TCACCAAACC GGCGAAGACC
6121  AGGTAAGCCT CTGAAGGACG C (SEQ ID NO:38)
```

**Figure 28**

```
   1  MIDTAPLAPP RAPRSNPIRD RVDWEAQRAA ALADPGAFHG AIARTVIHWY DPQHHCWIRF
  61  NESSQRWEGL DAATGAPVTV DYPADYQPWQ QAFDDSEAPF YRWFSGGLTN ACFNEVDRHV
 121  MMGYGDEVAY YFEGDRWDNS LNNGRGGPVV QETITRRRLL VEVVKAAQVL RDLGLKKGDR
 181  IALNMPNIMP QIYYTEAAKR LGILYTPVFG GFSDKTLSDR IHNAGARVVI TSDGAYRNAQ
 241  VVPYKEAYTD QALDKYIPVE TAQAIVAQTL ATLPLTESQR QTIITEVEAA LAGEITVERS
 301  DVMRGVGSAL AKLRDLDASV QAKVRTVLAQ ALVESPPRVE AVVVVRHTGQ EILWNEGRDR
 361  WSHDLLDAAL AKILANARAA GFDVHSENDL LNLPDDQLIR ALYASIPCEP VDAEYPMFII
 421  YTSGSTGKPK GVIHVHGGYV AGVVHTLRVS FDAEPGDTIY VIADPGWITG QSYMLTATMA
 481  GRLTGVIAEG SPLFPSAGRY ASIIERYGVQ IFKAGVTFLK TVMSNPQNVE DVRLYDMHSL
 541  RVATFCAEPV SPAVQQFGMQ IMTPQYINSY WATEHGGIVW THFYGNQDFP LRPDAHTYPL
 601  PWVMGDVWVA ETDESGTTRY RVADFDEKGE IVITAPYPYL TRTLWGDVPG FEAYLRGEIP
 661  LRAWKGDAER FVKTYWRRGP NGEWGYIQGD FAIKYPDGSF TLHGRPDDVI NVSGHRMGTE
 721  EIEGAILRDR QITPDSPVGN CIVVGAPHRE KGLTPVAFIQ PAPGRHLTGA DRRRLDELVR
 781  TEKGAVSVPE DYIEVSAFPE TRSGKYMRRF LRNMMLDEPL GDTTTLRNPE VLEEIAAKIA
 841  EWKRRQRMAE EQQIIERYRY FRIEYHPPTA SAGKLAVVTV TNPPVNALNE RALDELNTIV
 901  DHLARRQDVA AIVFTGQGAR SFVAGADIRQ LLEEIHTVEE AMALPNNAHL AFRKIERMNK
 961  PCIAAINGVA LGGGLEFAMA CHYRVADVYA EFGQPEINLR LLPGYGGTQR LPRLLYKRNN
1021  GTGLLRALEM ILGGRSVPAD EALKLGLIDA IATGDQDSLS LACALARAAI GADGQLIESA
1081  AVTQAFRHRH EQLDEWRKPD PRFADDELRS IIAHPRIERI IRQAHTVGRD AAVHRALDAI
1141  RYGIIHGFEA GLEHEAKLFA EAVVDPNGGK RGIREFLDRQ SAPLPTRRPL ITPEQEQLLR
1201  DQKELLPVGS PFFPGVDRIP KWQYAQAVIR DPDTGAAAHG DPIVAEKQII VPVERPRANQ
1261  ALIYVLASEV NFNDIWAITG IPVSRFDEHD RDWHVTGSGG IGLIVALGEE ARREGRLKVG
1321  DLVAIYSGQS DLLSPLMGLD PMAADFVIQG NDTPDGSHQQ FMLAQAPQCL PIPTDMSIEA
1381  AGSYILNLGT IYRALFTTLQ IKAGRTIFIE GAATGTGLDA ARSAARNGLR VIGMVSSSSR
1441  ASTLLAAGAH GAINRKDPEV ADCFTRVPED PSAWAAWEAA GQPLLAMFRA QNDGRLADYV
1501  VSHAGETAFP RSFQLLGEPR DGHIPTLTFY GATSGYHFTF LGKPGSASPT EMLRRANLRA
1561  GEAVLIYYGV GSDDLVDTGG LEAIEAARQM GARIVVVTVS DAQREFVLSL GFGAALRGVV
1621  SLAELKRRFG DEFEWPRTMP PLPNARQDPQ GLKEAVRRFN DLVFKPLGSA VGVFLRSADN
1681  PRGYPDLIIE RAAHDALAVS AMLIKPFTGR IVYFEDIGGR RYSFFAPQIW VRQRRIYMPT
1741  AQIFGTHLSN AYEILRLNDE ISAGLLTITE PAVVPWDELP EAHQAMWENR HTAATYVVNH
1801  ALPRLGLKNR DELYEAWTAG ER (SEQ ID NO:39)
```

**Figure 29**

ATGAGTGAAGAGTCTCTGGTTCTCAGCACAATTGAAGGCCCCATCGCCATCCTCACCCTC
AATCGCCCCCAGGCCCTCAATGCGCTCAGTCCGGCCTTGATTGATGACCTCATTCGCCAT
TTAGAAGCCTGCGATGCCGATGACACAATCCGCGTGATCATTATCACCGGCGCCGGACGG
GCATTTGCTGCCGGCGCTGATATCAAAGCGATGGCCAATGCCACGCCTATTGATATGCTC
ACCAGTGGCATGATTGCGCGCTGGGCACGCATCGCCGCGGTGCGCAAACCGGTGATTGCT
GCCGTGAATGGGTATGCGCTCGGTGGTGGTTGTGAATTGGCAATGATGTGCGACATCATC
ATCGCCAGTGAAAACGCGCAGTTCGGACAACCGGAAATCAATCTGGGCATCATTCCCGGT
GCTGGTGGCACCCAACGGCTGACCCGCGCCCTTGGCCCGTATCGCGCAATGGAATTGATC
CTGACCGGCGCGACCATCAGTGCTCAGGAAGCTCTCGCCCACGGCCTGGTGTGCCGGGTC
TGCCCGCCTGAAAGCCTGCTCGATGAAGCCCGTCGGATCGCGCAAACCATTGCCACCAAA
TCACCACTGGCTGTACAGTTGGCGAAAGAGGCAGTCCGTATGGCCGCCGAAACCACTGTG
CGCGAGGGGTTGGCTATCGAGCTGCGTAACTTCTATCTGCTGTTTGCCAGTGCTGACCAA
AAAGAGGGGATGCAGGCATTTATCGAGAAACGCGCTCCCAACTTCAGTGGTCGTTGA
(SEQ ID NO:40)

**Figure 30**

MSEESLVLSTIEGPIAILTLNRPQALNALSPALIDDLIRHLEACDADDTIRVIIITGAGR
AFAAGADIKAMANATPIDMLTSGMIARWARIAAVRKPVIAAVNGYALGGGCELAMMCDII
IASENAQFGQPEINLGIIPGAGGTQRLTRALGPYRAMELILTGATISAQEALAHGLVCRV
CPPESLLDEARRIAQTIATKSPLAVQLAKEAVRMAAETTVREGLAIELRNFYLLFASADQ
KEGMQAFIEKRAPNFSGR (SEQ ID NO:41)

**Figure 31**

GGCGTAATCCGACCGGCAGGTTAGGGTCTTCTACTGGGGTCAAGGCGCGTCTCCTTTTGG
TGGCGCGAGCAACCCGGCTTTTCCTGGCTTCAATGTACCATAGAGCGGTTACTTCGTGCA
ACGGGCGTGGTACAATCGAGAGCAACCTTTCGCAAAAGCTATCCAATCCTGCACACGTGC
ATCTGTTACAGGGTATTATTGTCGGCAAACGACAGTCCTGTCGTTTATGTACAAGGAGAT
CAACGTATGAGTGAAGAGTCTCTGGTTCTCAGCACAATTGAAGGCCCCATCGCCATCCTC
ACCCTCAATCGCCCCCAGGCCCTCAATGCGCTCAGTCCGGCCTTGATTGATGACCTCATT
CGCCATTTAGAAGCCTGCGATGCCGATGACACAATCCGCGTGATCATTATCACCGGCGCC
GGACGGGCATTTGCTGCCGGCGCTGATATCAAAGCGATGGCCAATGCCACGCCTATTGAT
ATGCTCACCAGTGGCATGATTGCGCGCTGGGCACGCATCGCCGCGGTGCGCAAACCGGTG
ATTGCTGCCGTGAATGGGTATGCGCTCGGTGGTGGTTGTGAATTGGCAATGATGTGCGAC
ATCATCATCGCCAGTGAAAACGCGCAGTTCGGACAACCGGAAATCAATCTGGGCATCATT
CCCGGTGCTGGTGGCACCCAACGGCTGACCCGCGCCCTTGGCCCGTATCGCGCAATGGAA
TTGATCCTGACCGGCGCGACCATCAGTGCTCAGGAAGCTCTCGCCCACGGCCTGGTGTGC
CGGGTCTGCCCGCCTGAAAGCCTGCTCGATGAAGCCCGTCGGATCGCGCAAACCATTGCC
ACCAAATCACCACTGGCTGTACAGTTGGCGAAAGAGGCAGTCCGTATGGCCGCCGAAACC
ACTGTGCGCGAGGGGTTGGCTATCGAGCTGCGTAACTTCTATCTGCTGTTTGCCAGTGCT
GACCAAAAAGAGGGGATGCAGGCATTTATCGAGAAACGCGCTCCCAACTTCAGTGGTCGT
TGATCACGCGCAGAACATGGCAGCAGGGGCAATACCTGCACGTACTGCCTCCTGCCGCCA
TACTACCAGATGATCGAGCAGTAAAGGGTAAATACTCTATCAATCTGGCCAGATAAGCGG
TTGGGTAACAACGCAATGCTCCAAAGGAGACGATCATGGACATACACGAGCGATTGCGAT
CTCTCGAACGCGAAAATGCT (SEQ ID NO:42)

**Figure 32**

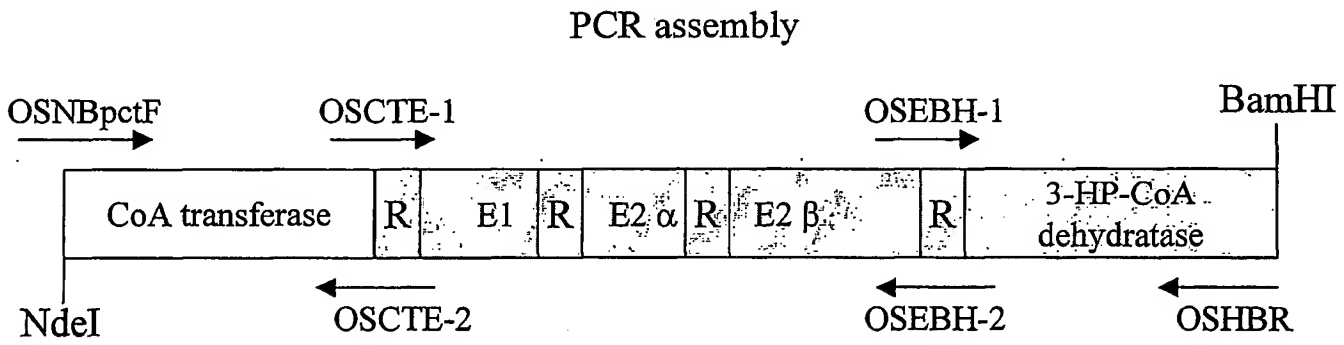
```
SEQ ID NO:40      1 ---------atgagtga----------agagt------------------
SEQ ID NO:43      1 ---------atgacgta----------cgaaa------------------
SEQ ID NO:44      1 atggccgccctgcgtgt----------cctgctgtcctgcgcccgcggcc
SEQ ID NO:45      1 atggcggccctgcgtgctctgctgcccagagc------------------

SEQ ID NO:40     14 ---------ct--------ctg---------gttctc-agcacaattgaa
SEQ ID NO:43     14 ---------cc--------atc---------ctggtcgagcgc----gat
SEQ ID NO:44     41 cgctgaggccc--------ccg---------gttcgc-tgtcccgcctgg
SEQ ID NO:45     33 ---------ctgcaactcgctgttgtccccagttcgc-tgcccagaattc

SEQ ID NO:40     37 ggccccatcgcc----------------------atcctcacc--------
SEQ ID NO:43     34 cagcgagttggc----------------------attatcacg--------
SEQ ID NO:44     73 cgtcccttcgcctcgggtgctaactttgagtacatcatcgcagaaaaaag
SEQ ID NO:45     73 cggcgcttcgcctcgggtgctaactttcagtacatcatcacg--------

SEQ ID NO:40     58 -------------------------------c------------------
SEQ ID NO:43     55 -------------------------------c------------------
SEQ ID NO:44    123 agggaagaataacaccgtggggttgatccaac------------------
SEQ ID NO:45    115 --------------------------------gaaaagaaaggaaagaata

SEQ ID NO:40     59 ----------------------tcaatcgccccaggccctcaatgcgctc
SEQ ID NO:43     56 ----------------------tgaaccgtccccaggcactgaacgcgctc
SEQ ID NO:44    155 ----------------------tgaaccgccccaaggccctcaatgcactt
SEQ ID NO:45    134 gcagcgtggggctgatccagttgaaccgtcccaaagcactcaatgcactt

SEQ ID NO:40     88 agtccggccttgattgatgacctcattc--gccatttagaagcctgcgat
SEQ ID NO:43     85 a--acagccagg--tgatgaacgaggtc--acca--gcgctgcaaccgaa
SEQ ID NO:44    184 tgcgatggcctgattgacgagctcaaccaggccctgaaga--tcttcgag
SEQ ID NO:45    184 tgcaatggactgattgaggagctcaacc--aagcactggagacctttgag

SEQ ID NO:40    136 ---gccgatgacaca---atccgcgtgatcattatcaccggcgccggacg
SEQ ID NO:43    127 ctggacgatgacccggacattggggcgatcatcatcaccggttcggccaa
SEQ ID NO:44    232 ---gaggacccggcc---gttgggggcattgtcctcaccggcggggataa
SEQ ID NO:45    232 ---gaagatcccgct---gtgggcgccattgtgctcactggtggggagaa

SEQ ID NO:40    180 ggcatttgctgccggcgctgatatcaaagcgatggccaa-------tgcc
SEQ ID NO:43    177 agcgtttgccgccggagccgacatcaaagaaatggccga-------cctg
SEQ ID NO:44    276 ggcctttgcagctggagctgatatcaaggaaatgcagaacctgagtttcc
SEQ ID NO:45    276 ggcctttgcagccggagctgacatcaaggaaatgcagaa-------ccgg

SEQ ID NO:40    223 acgcctattgatatgctcaccagtggcatgattgcgcgc---tgggcacg
SEQ ID NO:43    220 acgttcgccgacgcgttcaccgccgacttcttcgccacc---tggggcaa
SEQ ID NO:44    326 aggactgtt-------actccagcaagttcttgaagcac---tggggcca
SEQ ID NO:45    319 acatttcagga-ctgttactca--ggcaagttcctgagccactgggacca

SEQ ID NO:40    270 catcgccgcggtgcgcaaaccggtgattgctgccgtgaatgggtatgcgc
SEQ ID NO:43    267 gctggccgccgtgcgcaccccgacgatcgccgcggtggcgggatacgcgc
SEQ ID NO:44    366 cctcacccaggtcaagaagccagtcatcgctgctgtcaatggctatccgt
SEQ ID NO:45    366 tatcacccggatcaagaaaccggtcatcgcggctgtcaatggctatgctc

SEQ ID NO:40    320 tcggtggtggttgtgaattggcaatgatgtgcgacatcatcatcgccagt
SEQ ID NO:43    317 tcggcggtggctgcgagctggcgatgatgtgcgacgtgctgatcgccgcc
SEQ ID NO:44    416 ttggcgggggctgtgagcttgccatgatgtgtgatatcatctatgccggt
SEQ ID NO:45    416 ttggtgggggctgtgaacttgccatgatgtgcgatatcatctatgctggt
```

```
SEQ ID NO:40    370 gaaaacgcgcagttcggacaaccggaaatcaatctgggcatcattcccgg
SEQ ID NO:43    367 gacaccgcgaagttcggacagcccgagataaagctgggcgtgctgccagg
SEQ ID NO:44    466 gagaaggcccagtttgcacagccggagatcttaataggaaccatcccagg
SEQ ID NO:45    466 gagaaagcccagtttggacagccagaaatcctcctggggaccatcccagg

SEQ ID NO:40    420 tgctggtggcacccaacggctgacccgcgcccttggcccgtatcgcgcaa
SEQ ID NO:43    417 catgggcggctcccagcggctgacccgggctatcggcaaggctaaggcga
SEQ ID NO:44    516 tgcaggcggcacccagagactcacccgtgctgttgggaagtcgctggagc
SEQ ID NO:45    516 tgcagggggcactcagagactcacccgagcagtcggcaaatcactagcaa

SEQ ID NO:40    470 tggaattgatcctgaccggcgcgaccatcagtgctcaggaagctctcgcc
SEQ ID NO:43    467 tggacctcatcctgaccgggcgcaccatggacgccgccgaggc-cgagcg
SEQ ID NO:44    566 tggagatggtcctcaccggtgacgcgatctcagcccaggacgc-caagca
SEQ ID NO:45    566 tggagatggtcctcactggtgaccgaatttcagcacaggatgc-caagca

SEQ ID NO:40    520 ca-c-ggcctggtgtgccgggtctgcccgcctgaaagcctgctcgatgaa
SEQ ID NO:43    516 cagc-ggtctggtttcacgggtggtgccggccgacgacttgctgaccgaa
SEQ ID NO:44    615 ag-caggtcttgtcagcaagatttgtcctgttgagacactggtggaagaa
SEQ ID NO:45    615 ag-caggtcttgtaagcaagatttttcccgttgaaacactggttgaagag

SEQ ID NO:40    568 gcccgtcggatcgcgcaaaccattgccaccaaatcaccactggctgtaca
SEQ ID NO:43    565 gccagggccactgccacgaccatttcgcagatgtcggcctcggcggcccg
SEQ ID NO:44    664 gccatccagtgtgcagaaaaaattgccagcaattctaaaattgtagtagc
SEQ ID NO:45    664 gccatccaatgtgcagaaaagatcgccaacaattccaagatcatagtagc

SEQ ID NO:40    618 gttggcgaaagaggcagtccgtatggccgccgaaaccactgtgcgcgagg
SEQ ID NO:43    615 gatggccaaggaggccgtcaaccgggctttcgaatccagtttgtccgagg
SEQ ID NO:44    714 gatggccaaagaatcagtgaatgcagcttttgaaatgacattaacagaag
SEQ ID NO:45    714 catggcgaaagaatctgtgaatgcagcctttgaaatgacgttaacagaag

SEQ ID NO:40    668 ggttggctatcgagctgcgtaacttctatctgctgtttgccagtgctgac
SEQ ID NO:43    665 ggctgctctacgaacgccggctttccattcggctttcgcgaccgaagac
SEQ ID NO:44    764 gaagtaagttggagaagaaactcttttattcaacctttgccactgatgac
SEQ ID NO:45    764 gaaataagctggagaagaagctcttctattccacctttgccactgatgac

SEQ ID NO:40    718 caaaaagaggggatgcaggcatttatcgagaaacgcgctcccaacttcag
SEQ ID NO:43    715 caatccgaaggtatggcagcgttcatcgagaaacgcgctccccagttcac
SEQ ID NO:44    814 cggaaagaagggatgaccgcgtttgtggaaaagagaaaggccaacttcaa
SEQ ID NO:45    814 cggagagaagggatgtctgcctttgtggagaaaaggaaggccaacttcaa

SEQ ID NO:40    768 tggtcgttga
SEQ ID NO:43    765 ccaccgatga
SEQ ID NO:44    864 agaccagtga
SEQ ID NO:45    864 agaccactga
```

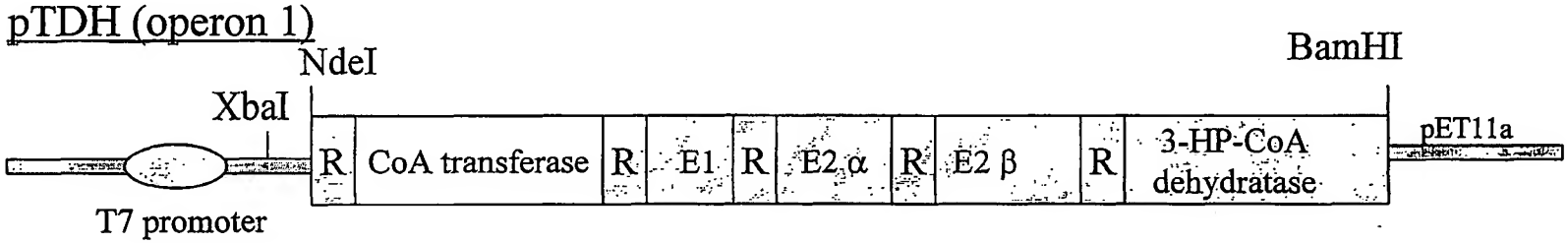
**Figure 33**

```
SEQ ID NO:41      1 -mseeslv-------lstiegp----------------------------
SEQ ID NO:46      1 -mtyetil-------ver-dqr----------------------------
SEQ ID NO:47      1 -maalrvl-------lscargplrppvrcpawrpfasganfeyiiaekrg
SEQ ID NO:48      1 maalrallpracnsllspvrcpefrrfasganfqyiitekkgknss----

SEQ ID NO:41     15 ----iailtlnrpqalnalspaliddlirhleacdaddtirviiitgagr
SEQ ID NO:46     14 ----vgiitlnrpqalnalnsqvmnevtsaateldddpdigaiiitgsak
SEQ ID NO:47     43 knntvgliqlnrpkalnalcdglidelnqalkifeedpavgaivltggdk
SEQ ID NO:48     47 ----vgliqlnrpkalnalcnglieelnqaletfeedpavgaivltggek

SEQ ID NO:41     61 afaagadikamanatpidmltsgmiarwariaavrkpviaavngyalggg
SEQ ID NO:46     60 afaagadikemadltfadaftadffatwgklaavrtptiaavagyalggg
SEQ ID NO:47     93 afaagadikemqnlsfqdcysskflkhwdhltqvkkpviaavngyafggg
SEQ ID NO:48     93 afaagadikemqnrtfqdcysgkflshwdhitrikkpviaavngyalggg

SEQ ID NO:41    111 celammcdiiiasenaqfgqpeinlgiipgaggtqrltralgpyrameli
SEQ ID NO:46    110 celammcdvliaadtakfgqpeiklgvlpgmggsqrltraigkakamdli
SEQ ID NO:47    143 celammcdiiyagekaqfaqpeiligtipgaggtqrltravgkslamemv
SEQ ID NO:48    143 celammcdiiyagekaqfgqpeillgtipgaggtqrltravgkslamemv

SEQ ID NO:41    161 ltgatisaqealahglvcrvcppeslldearriaqtiatksplavqlake
SEQ ID NO:46    160 ltgrtmdaaeaersglvsrvvpaddlltearatattisqmsasaarmake
SEQ ID NO:47    193 ltgdrisaqdakqaglvskicpvetlveeaiqcaekiasnskivvamake
SEQ ID NO:48    193 ltgdrisaqdakqaglvskifpvetlveeaiqcaekiannskiivamake

SEQ ID NO:41    211 avrmaaettvreglaielrnfyllfasadqkegmqafiekrapnfsgr
SEQ ID NO:46    210 avnrafesslseglleyerrlfhsafatedqsegmaafiekrapqfthr
SEQ ID NO:47    243 svnaafemtltegsklekklfystfatddrkegmtafvekrkanfkdq
SEQ ID NO:48    243 svnaafemtltegnklekklfystfatddrregmsafvekrkanfkdh
```

# Figure 34

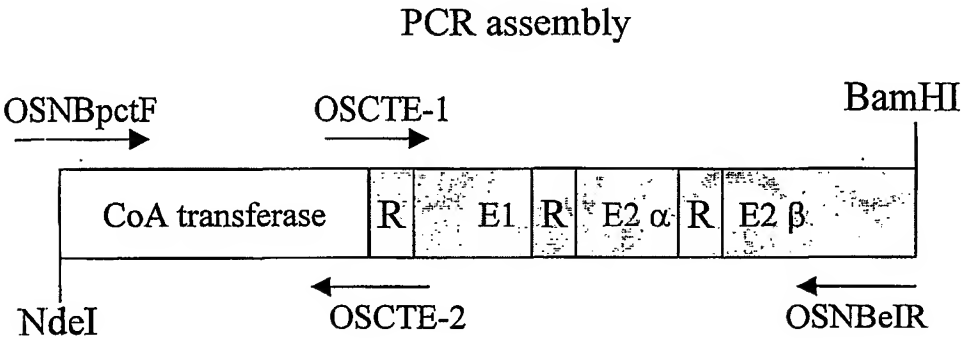



Cloning of PCR assembly in pET11a vector

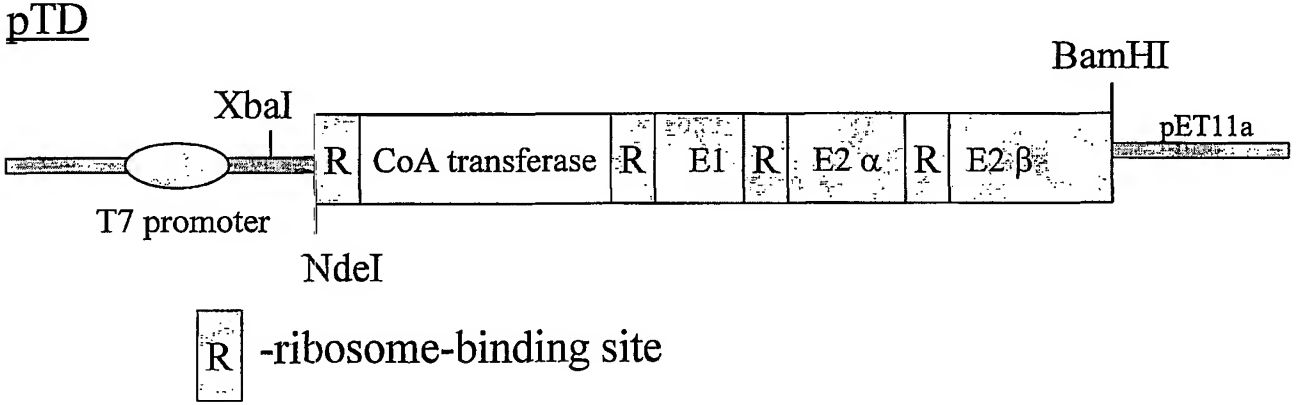



R -ribosome-binding site

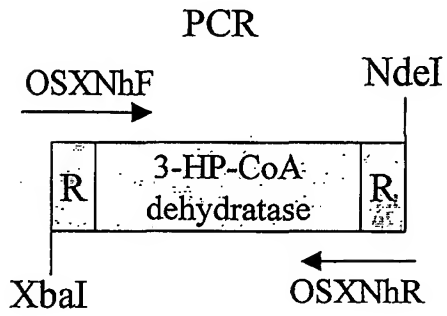
# Figure 35A




Cloning of PCR assembly in pET11a vector



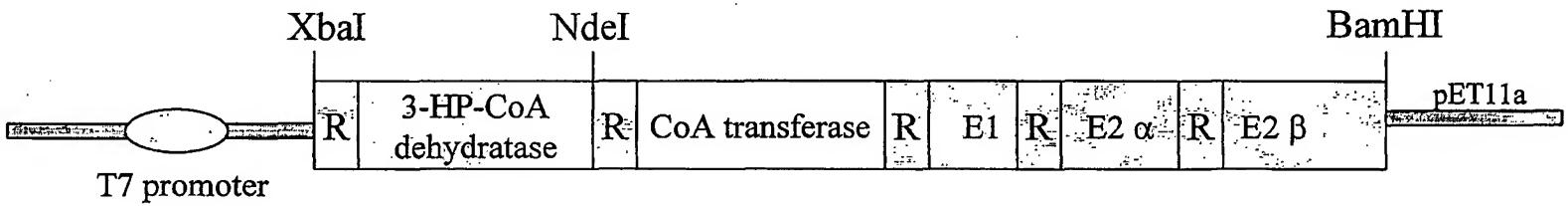
# Figure 35B




Cloning of 3-hydroxypropionyl-CoA dehydratase in pTD

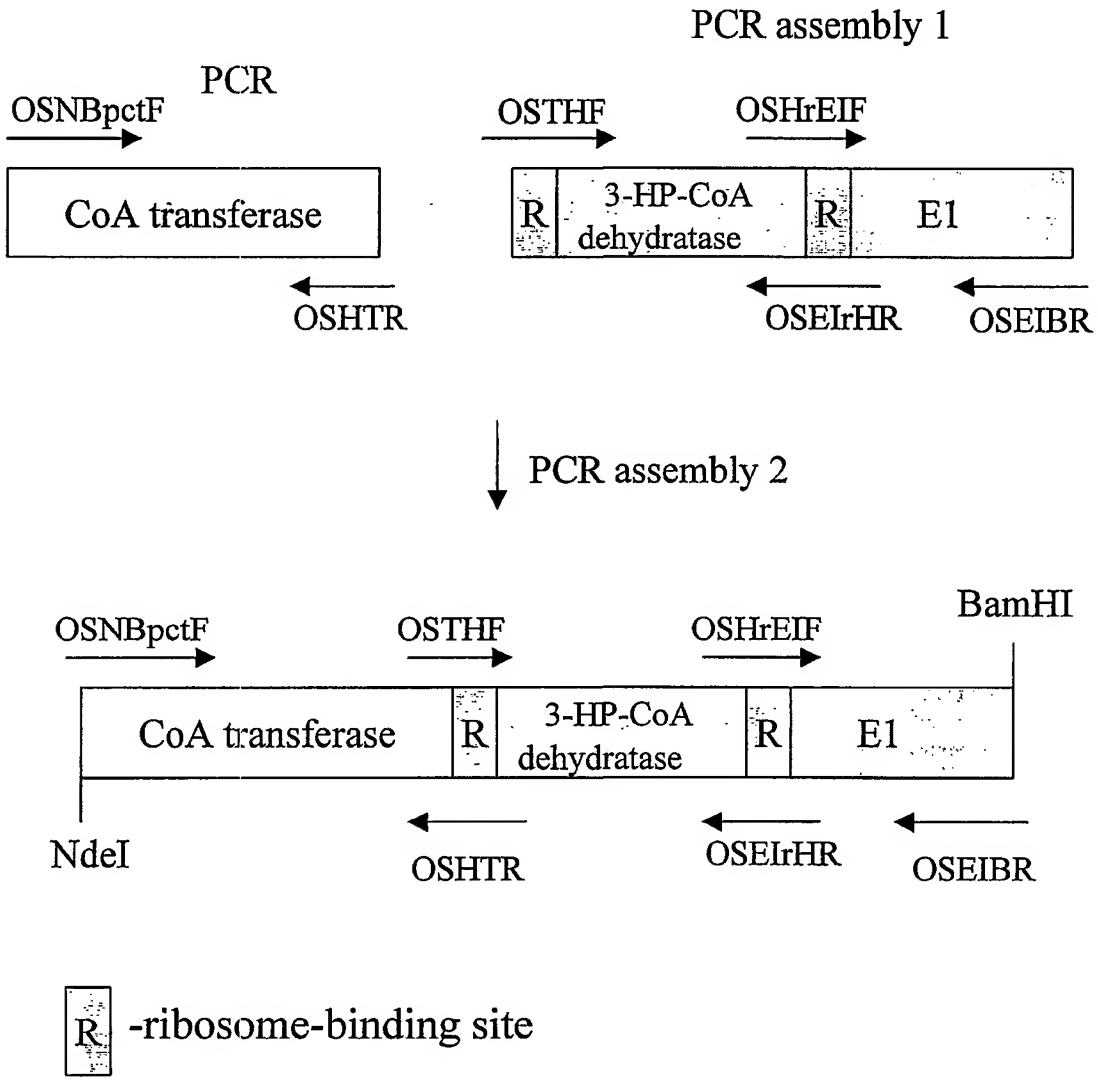


pHTD (operon 2)



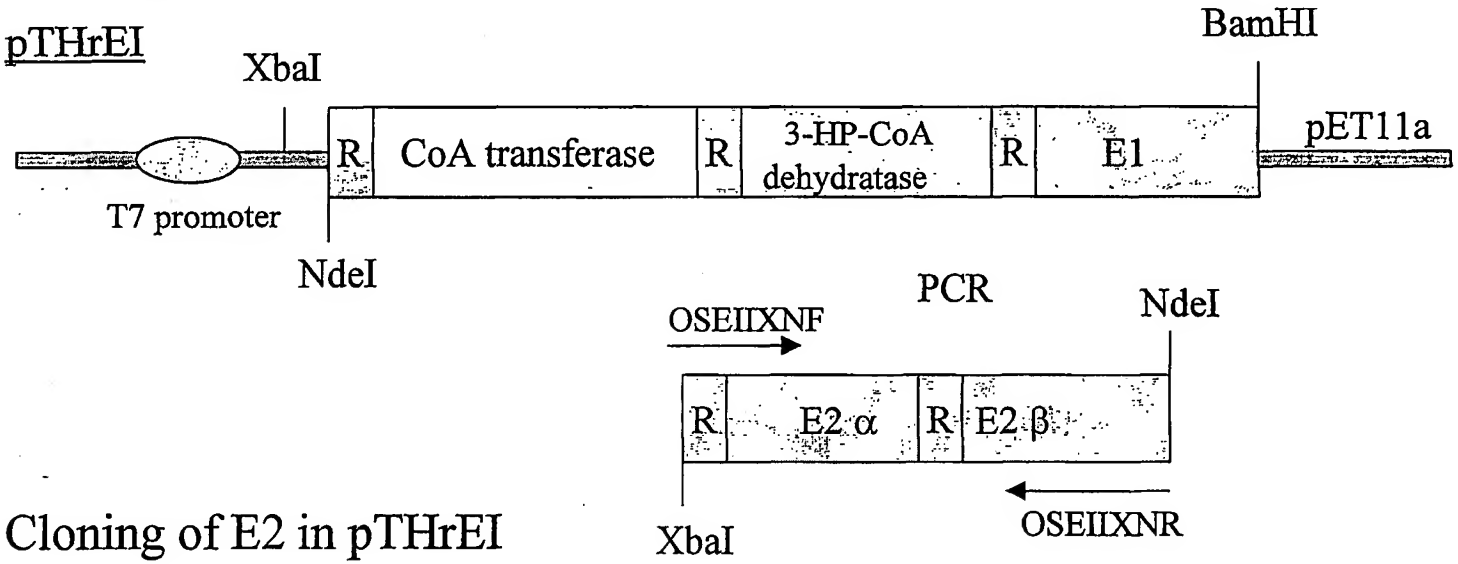

R -ribosome-binding site

# Figure 36A

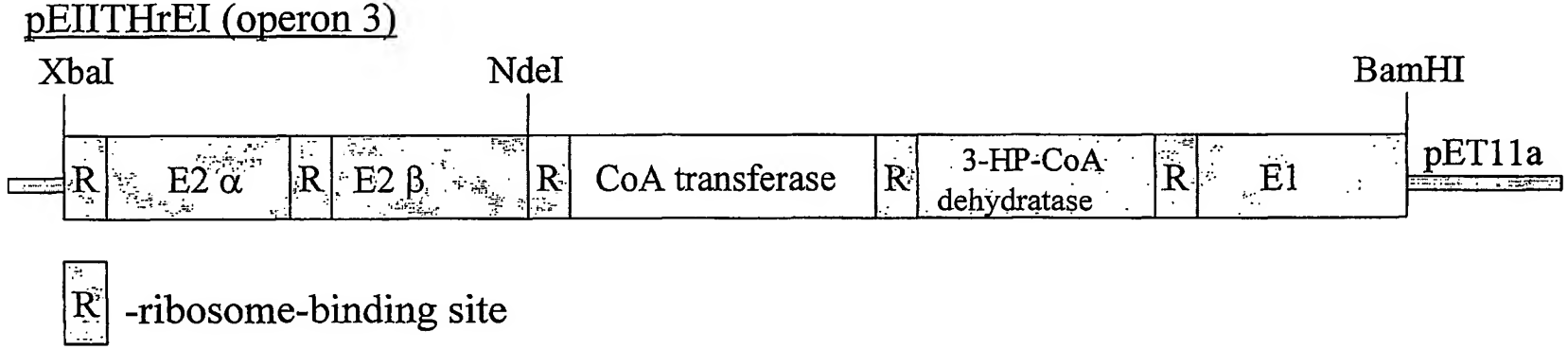

PCR assembly 1
PCR
OSNBpctF
OSTHF
OSHrEIF
CoA transferase
R
3-HP-CoA dehydratase
R
E1
OSHTR
OSEIrHR
OSEIBR
PCR assembly 2
BamHI
OSNBpctF
OSTHF
OSHrEIF
CoA transferase
R
3-HP-CoA dehydratase
R
E1
NdeI
OSHTR
OSEIrHR
OSEIBR
R -ribosome-binding site

# Figure 36B

## Cloning of PCR assembly 2 in pET11a vector

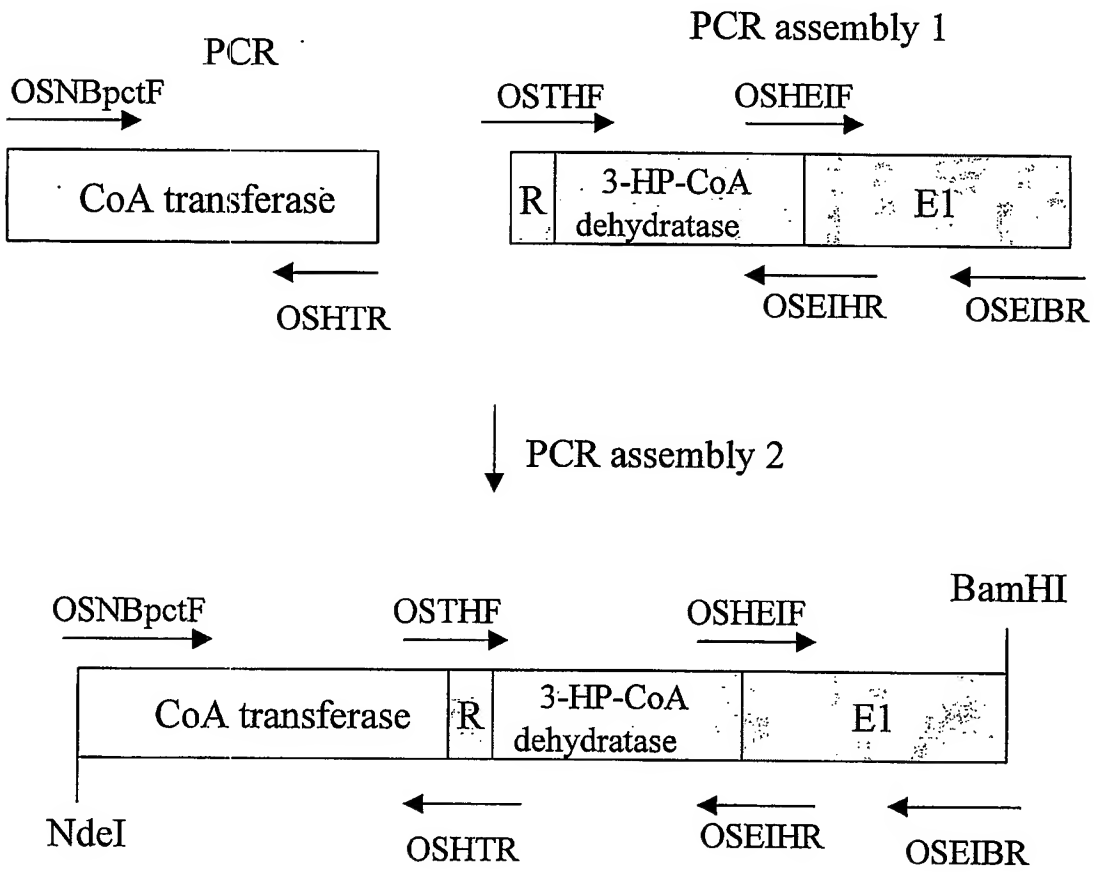



## Cloning of E2 in pTHrEI



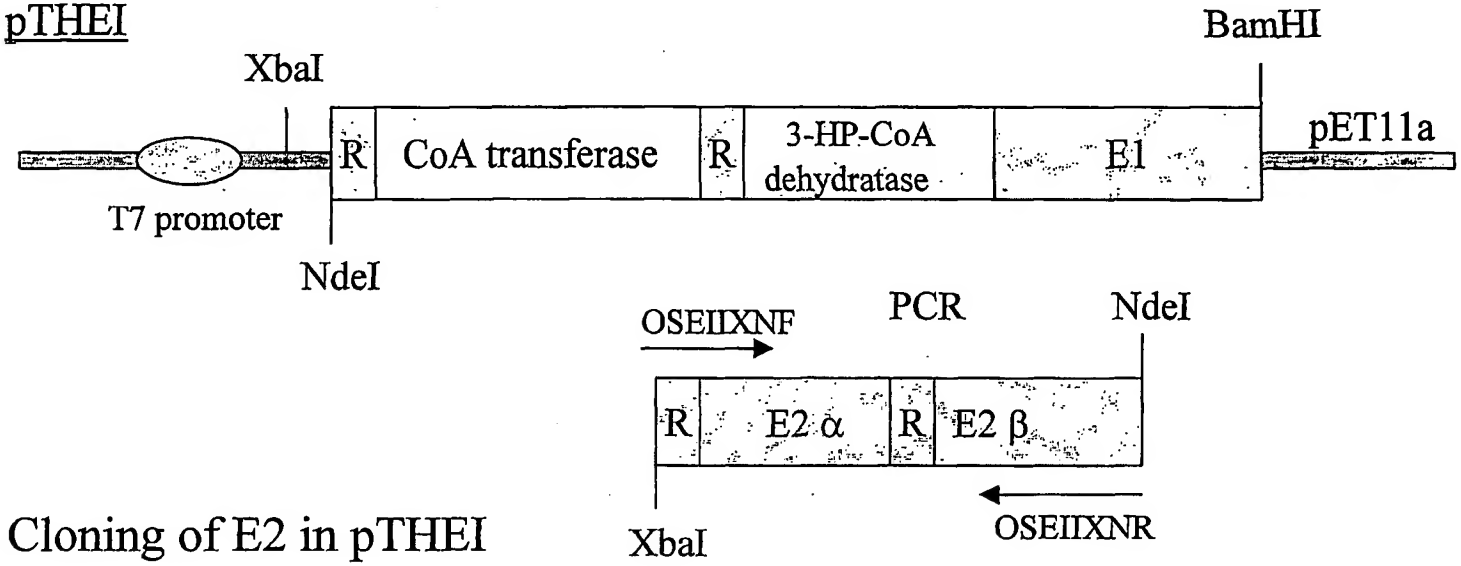

R -ribosome-binding site

# Figure 37A


PCR
PCR assembly 1
OSNBpctF
OSTHF
OSHEIF
CoA transferase
R
3-HP-CoA dehydratase
E1
OSHTR
OSEIHR
OSEIBR
PCR assembly 2
OSNBpctF
OSTHF
OSHEIF
BamHI
CoA transferase
R
3-HP-CoA dehydratase
E1
NdeI
OSHTR
OSEIHR
OSEIBR
R -ribosome-binding site

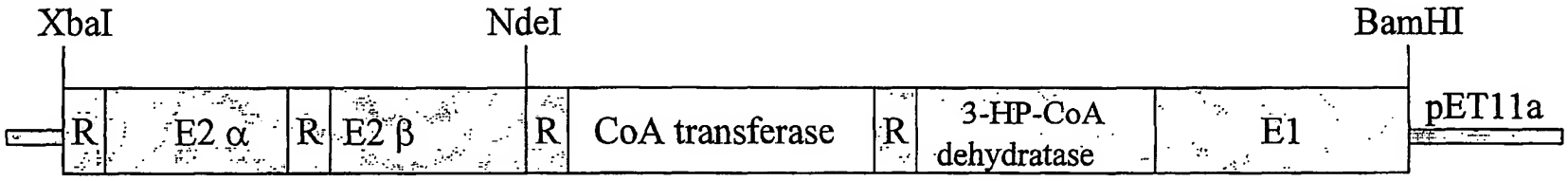
# Figure 37B

## Cloning of PCR assembly 2 in pET11a vector



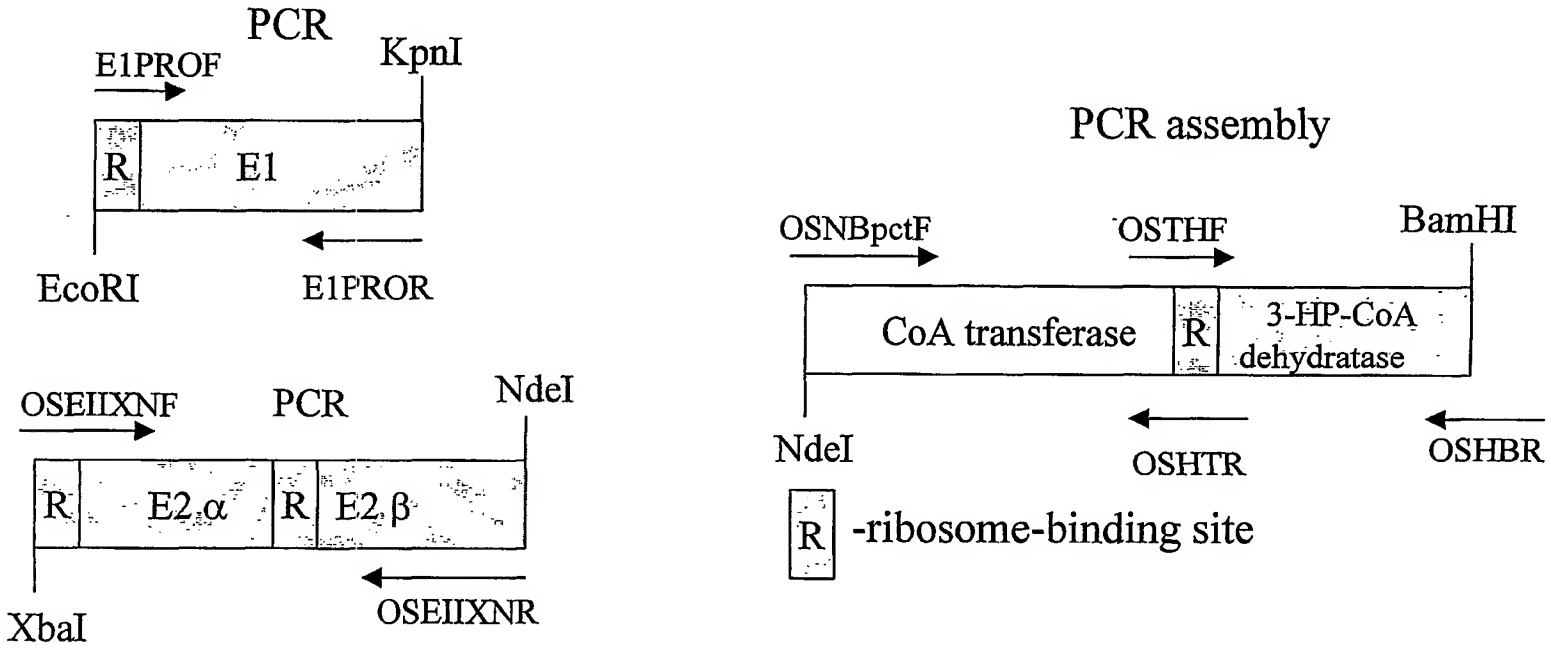

## Cloning of E2 in pTHEI

pEIITHEI (operon 4)

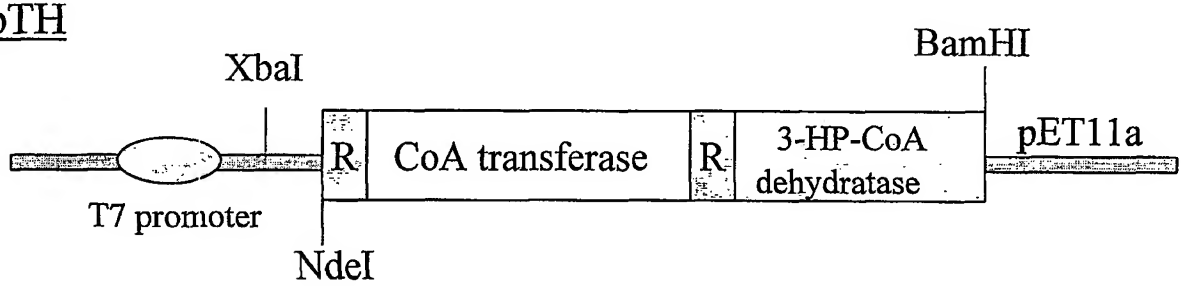



R -ribosome-binding site

# Figure 38A

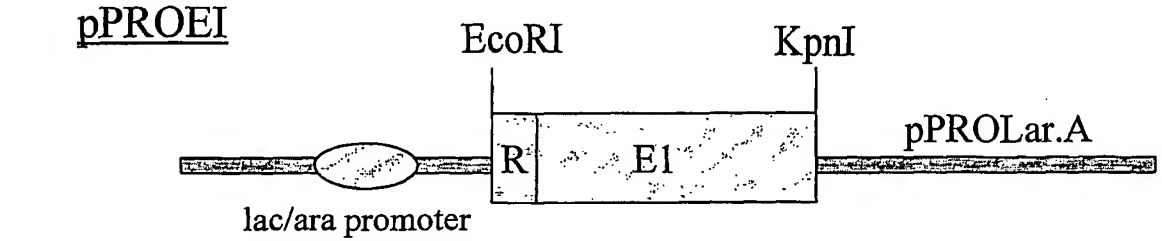



Cloning of PCR assembly in pET11a vector



# Figure 38B

## Cloning of E1 gene separately in pPROLar.A vector




## Cloning of E2 in pTH

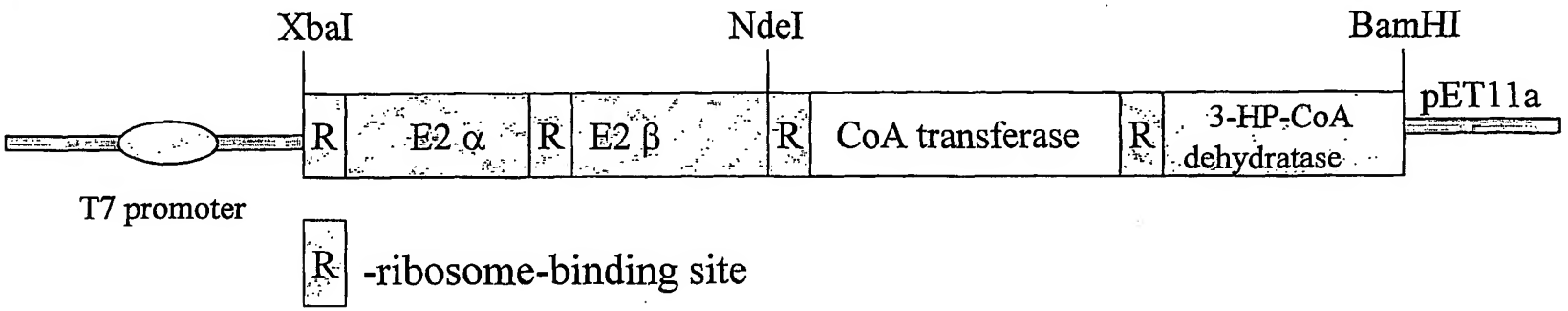


Figure 39

ATGATCGACACTGCGCCCCTTGCCCCACCACGGGCGCCCCGCTCTAATCCGATTCGGGAT
CGAGTTGATTGGGAAGCTCAGCGCGCTGCTGCGCTGGCAGATCCCGGTGCCTTTCATGGC
GCGATTGCCCGGACAGTTATCCACTGGTACGACCCACAACACCATTGCTGGATTCGCTTC
AACGAGTCTAGTCAGCGTTGGGAAGGGCTGGATGCCGCTACCGGTGCCCCTGTAACGGTA
GACTATCCCGCCGATTATCAGCCCTGGCAACAGGCGTTTGATGATAGTGAAGCGCCGTTT
TACCGCTGGTTTAGTGGTGGGTTGACAAATGCCTGCTTTAATGAAGTAGACCGGCATGTC
ATGATGGGCTATGGCGACGAGGTGGCCTACTACTTTGAAGGTGACCGCTGGGATAACTCG
CTCAACAATGGTCGTGGTGGTCCGGTTGTCCAGGAGACAATCACGCGGCGGCGCCTGTTG
GTGGAGGTGGTGAAGGCTGCGCAGGTGTTGCGTGATCTGGGCCTGAAGAAGGGTGATCGG
ATTGCTCTGAATATGCCGAATATTATGCCGCAGATTTATTATACGGAAGCGGCAAAACGA
CTGGGTATTCTGTACACGCCGGTCTTCGGTGGCTTCTCGGACAAGACTCTTTCCGACCGT
ATTCACAATGCCGGTGCACGAGTGGTGATTACCTCTGATGGTGCGTACCGCAACGCGCAG
GTGGTGCCCTACAAAGAAGCGTATACCGATCAGGCGCTCGATAAGTATATTCCGGTTGAG
ACGGCGCAGGCGATTGTTGCGCAGACCCTGGCCACCTTGCCCCTGACTGAGTCGCAGCGC
CAGACGATCATCACCGAAGTGGAGGCCGCACTGGCCGGTGAGATTACGGTTGAGCGCTCG
GACGTGATGCGTGGGGTTGGTTCTGCCCTCGCAAAGCTCCGCGATCTTGATGCAAGCGTG
CAGGCAAAGGTGCGTACAGTACTGGCGCAGGCGCTGGTCGAGTCGCCGCCGCGGGTTGAA
GCTGTGGTGGTTGTGCGTCATACCGGTCAGGAGATTTTGTGGAACGAGGGGCGAGATCGC
TGGAGTCACGACTTGCTGGATGCTGCGCTGGCGAAGATTCTGGCCAATGCGCGTGCTGCC
GGCTTTGATGTGCACAGTGAGAATGATCTGCTCAATCTCCCCGATGACCAGCTTATCCGT
GCGCTCTACGCCAGTATTCCCTGTGAACCGGTTGATGCTGAATATCCGATGTTTATCATT
TACACATCGGGTAGCACCGGTAAGCCCAAGGGTGTGATCCACGTTCACGGCGGTTATGTC
GCCGGTGTGGTGCACACCTTGCGGGTCAGTTTTGACGCCGAGCCGGGTGATACGATATAT
GTGATCGCCGATCCGGGCTGGATCACCGGTCAGAGCTATATGCTCACAGCCACAATGGCC
GGTCGGCTGACCGGGGTGATTGCCGAGGGATCACCGCTCTTCCCCTCAGCCGGGCGTTAT
GCCAGCATCATCGAGCGCTATGGGGTGCAGATCTTTAAGGCGGGTGTGACCTTCCTCAAG
ACAGTGATGTCCAATCCGCAGAATGTTGAAGATGTGCGACTCTATGATATGCACTCGCTG
CGGGTTGCAACCTTCTGCGCCGAGCCGGTCAGTCCGGCGGTGCAGCAGTTTGGTATGCAG
ATCATGACCCCGCAGTATATCAATTCGTACTGGGCGACCGAGCACGGTGGAATTGTCTGG
ACGCATTTCTACGGTAATCAGGACTTCCCGCTTCGTCCCGATGCCCATACCTATCCCTTG
CCCTGGGTGATGGGTGATGTCTGGGTGGCCGAAACTGATGAGAGCGGGACGACGCGCTAT
CGGGTCGCTGATTTCGATGAGAAGGGCGAGATTGTGATTACCGCCCCGTATCCCTACCTG
ACCCGCACACTCTGGGGTGATGTGCCCGGTTTCGAGGCGTACCTGCGCGGTGAGATTCCG
CTGCGGGCCTGGAAGGGTGATGCCGAGCGTTTCGTCAAGACCTACTGGCGACGTGGGCCA
AACGGTGAATGGGGCTATATCCAGGGTGATTTTGCCATCAAGTACCCCGATGGTAGCTTC
ACGCTCCACGGACGCCCTGACGATGTGATCAATGTGTCGGGCCACCGTATGGGCACCGAG
GAGATTGAGGGTGCCATTTTGCGTGACCGCCAGATCACGCCCGACTCGCCCGTCGGTAAT
TGTATTGTGGTCGGTGCGCCGCACCGTGAGAAGGGTCTGACCCCGGTTGCCTTCATTCAA
CCTGCGCCTGGCCGTCATCTGACCGGCGCCGACCGGCGCCGTCTCGATGAGCTGGTGCGT
ACCGAGAAGGGGGCGGTCAGTGTCCCAGAGGATTACATCGAGGTCAGTGCCTTTCCCGAA
ACCCGCAGCGGGAAGTATATGCGGCGCTTTTTGCGCAATATGATGCTCGATGAACCACTG
GGTGATACGACGACGTTGCGCAATCCTGAAGTGCTCGAAGAGATTGCAGCCAAGATCGCT
GAGTGGAAACGCCGTCAGCGTATGGCCGAAGAGCAGCAGATCATCGAACGCTATCGCTAC
TTCCGGATCGAGTATCACCCACCAACGGCCAGTGCGGGTAAACTCGCGGTAGTGACGGTG
ACAAATCCGCCGGTGAACGCACTGAATGAGCGTGCGCTCGATGAGTTGAACACAATTGTT
GACCACCTGGCCCGTCGTCAGGATGTTGCCGCAATTGTCTTCACCGGACAGGGCGCCAGG
AGTTTTGTCGCCGGCGCTGATATTCGCCAGTTGCTCGAAGAGATTCATACGGTTGAAGAG
GCAATGGCCCTGCCGAATAACGCCCATCTTGCTTTCCGCAAGATTGAGCGTATGAATAAG

CCGTGTATCGCGGCGATCAACGGTGTGGCGCTCGGTGGTGGTCTGGAATTCGCCATGGCC
TGCCATTACCGGGTTGCCGATGTCTATGCCGAATTCGGTCAGCCAGAGATTAATCTGCGC
TTGCTACCTGGTTATGGTGGCACGCAGCGCTTGCCGCGCCTGTTGTACAAGCGCAACAAC
GGCACCGGTCTGCTCCGAGCGCTGGAGATGATTCTGGGTGGGCGTAGCGTACCGGCTGAT
GAGGCGCTGAAGCTGGGTCTGATCGATGCCATTGCTACCGGCGATCAGGACTCACTGTCG
CTGGCATGCGCGTTAGCCCGTGCCGCAATCGGCGCCGATGGTCAGTTGATCGAGTCGGCT
GCGGTGACCCAGGCTTTCCGCCATCGCCACGAGCAGCTTGACGAGTGGCGCAAACCAGAC
CCGCGCTTTGCCGATGACGAACTGCGCTCGATTATCGCCCATCCACGTATCGAGCGGATT
ATCCGGCAGGCCCATACCGTTGGGCGCGATGCGGCAGTGCATCGGGCACTGGATGCAATC
CGCTATGGCATTATCCACGGCTTCGAGGCCGGTCTGGAGCACGAGGCGAAGCTCTTTGCC
GAGGCAGTGGTTGACCCGAACGGTGGCAAGCGTGGTATTCGCGAGTTCCTCGACCGCCAG
AGTGCGCCGTTGCCAACCCGCCGACCATTGATTACACCTGAACAGGAGCAACTCTTGCGC
GATCAGAAAGAACTGTTGCCGGTTGGTTCACCCTTCTTCCCCGGTGTTGACCGGATTCCG
AAGTGGCAGTACGCGCAGGCGGTTATTCGTGATCCGGACACCGGTGCGGCGGCTCACGGC
GATCCCATCGTGGCTGAAAAGCAGATTATTGTGCCGGTGGAACGCCCCCGCGCCAATCAG
GCGCTGATCTATGTTCTGGCCTCGGAGGTGAACTTCAACGATATCTGGGCGATTACCGGT
ATTCCGGTGTCACGGTTTGATGAGCACGACCGCGACTGGCACGTTACCGGTTCAGGTGGC
ATCGGCCTGATCGTTGCGCTGGGTGAAGAGGCGCGACGCGAAGGCCGGCTGAAGGTGGGT
GATCTGGTGGCGATCTACTCCGGGCAGTCGGATCTGCTCTCACCGCTGATGGGCCTTGAT
CCGATGGCCGCCGATTTCGTCATCCAGGGGAACGACACGCCAGATGGATCGCATCAGCAA
TTTATGCTGGCCCAGGCCCCGCAGTGTCTGCCCATCCCAACCGATATGTCTATCGAGGCA
GCCGGCAGCTACATCCTCAATCTCGGTACGATCTATCGCGCCCTCTTTACGACGTTGCAA
ATCAAGGCCGGACGCACCATCTTTATCGAGGGTGCGGCGACCGGTACCGGTCTGGACGCA
GCGCGCTCGGCGGCCCGGAATGGTCTGCGCGTAATTGGAATGGTCAGTTCGTCGTCACGT
GCGTCTACGCTGCTGGCTGCGGGTGCCCACGGTGCGATTAACCGTAAAGACCCGGAGGTT
GCCGATTGTTTCACGCGCGTGCCCGAAGATCCATCAGCCTGGGCAGCCTGGGAAGCCGCC
GGTCAGCCGTTGCTGGCGATGTTCCGGGCGCAGAACGACGGGCGACTGGCCGATTATGTG
GTCTCGCACGCGGGCGAGACGGCCTTCCCGCGCAGTTTCCAGCTTCTCGGCGAGCCACGC
GATGGTCACATTCCGACGCTCACATTCTACGGTGCCACCAGTGGCTACCACTTCACCTTC
CTGGGTAAGCCAGGGTCAGCTTCGCCGACCGAGATGCTGCGGCGGGCCAATCTCCGCGCC
GGTGAGGCGGTGTTGATCTACTACGGGGTTGGGAGCGATGACCTGGTAGATACCGGCGGT
CTGGAGGCTATCGAGGCGGCGCGGCAAATGGGAGCGCGGATCGTCGTCGTTACCGTCAGC
GATGCGCAACGCGAGTTTGTCCTCTCGTTGGGCTTCGGGGCTGCCCTACGTGGTGTCGTC
AGCCTGGCGGAACTCAAACGGCGCTTCGGCGATGAGTTTGAGTGGCCGCGCACGATGCCG
CCGTTGCCGAACGCCCGCCAGGACCCGCAGGGTCTGAAAGAGGCTGTCCGCCGCTTCAAC
GATCTGGTCTTCAAGCCGCTAGGAAGCGCGGTCGGTGTCTTCTTGCGGAGTGCCGACAAT
CCGCGTGGCTACCCCGATCTGATCATCGAGCGGGCTGCCCACGATGCACTGGCGGTGAGC
GCGATGCTGATCAAGCCCTTCACCGGACGGATTGTCTACTTCGAGGACATTGGTGGGCGG
CGTTACTCCTTCTTCGCACCGCAAATCTGGGTGCGCCAGCGCCGCATCTACATGCCGACG
GCACAGATCTTTGGTACGCACCTCTCAAATGCGTATGAAATTCTGCGTCTGAATGATGAG
ATCAGCGCCGGTCTGCTGACGATTACCGAGCCGGCAGTGGTGCCGTGGGATGAACTACCC
GAAGCACATCAGGCGATGTGGGAAAATCGCCACACGGCGGCCACTTATGTGGTGAATCAT
GCCTTACCACGTCTCGGCCTAAAGAACAGGGACGAGCTGTACGAGGCGTGGACGGCCGGC
GAGCGGTAG (SEQ ID NO:129)

**Figure 40**

```
SEQ ID NO:39      1 ------------------------midtaplappraprsnpirdrvdwe
SEQ ID NO:130     1 mglpeervrsgsgsrgqeeagaggrarswsp--ppevsrsahvpslqryr
SEQ ID NO:131     1 ----------------------------mslelkekeselpfdeqiind
                                                 PL PP   RS P

SEQ ID NO:39     26 aqraaaladpgafhgaiartvihwydpqhhcwirfnessqrwegldaatg
SEQ ID NO:130    49 elhrrsveeprefwgdiake-fywktpcpgpflryn--------------
SEQ ID NO:131    22 kwrs------------kytpidayfkfhrqtvenlesf--wesv------
                      R      P  F G IA T I WY P H    R NES   WE

SEQ ID NO:39     76 apvtvdypadyqpwqqafddseap-fyrwfsggltnacfnevdrhvm-mg
SEQ ID NO:130    84 -----------------fdvtkgkifiewmkgattnicynvldrnvhekk
SEQ ID NO:131    52 -akelew---fkpwdkvldasnpp-fykwfvggrlnlsylavdrhvk-tw
                              PW   FD S  P FY WF GG TN C N VDRHV

SEQ ID NO:39    124 ygdevayyfegdrwdnslnngrggpvvqetitrrrllvevvkaaqvlr-d
SEQ ID NO:130   117 lgdkvafywegne-----------pgettqityhqllvqvcqfsnvlr-k
SEQ ID NO:131    96 rknklaiewegepvden-----gyptdrrkltyydlyrevnrvaymlkqn
                     GD VA Y EG   D       G P     IT   LLVEV   A VLR

SEQ ID NO:39    173 lglkkgdrialnmpnimpqiyyte-aakrlgilytpvfggfsdktlsdri
SEQ ID NO:130   155 qgiqkgdrvaiympmipelvvaml-acarigalhsivfagfsseslceri
SEQ ID NO:131   141 fgvkkgdkitlylp-mvpeipitmlaawrigaitsvvfsgfsadalaeri
                     G KKGDRIAL MP I P    T  AA R G L   VF GFS   L  RI

SEQ ID NO:39    222 hnagarvvitsdgayrnaqvvpykeaytdqal----dkyipvetaqaiva
SEQ ID NO:130   204 ldsscsllittdafyrgeklvnlkel-adealqkcqekgfpvrc--civv
SEQ ID NO:131   190 ndsqsrivitadgfwrrgrvvrlkev-----------------------
                         R VIT DG YR   VV KE   D AL     K  PV     IV

SEQ ID NO:39    268 qtlatlpltesqrqtiiteveaalageitversdvmrgvgsalaklrdld
SEQ ID NO:130   251 khlgrael-----------------------------gmgdsts------
SEQ ID NO:131   216 --------------------vdaal-------------------------
                      L    L           V AAL             G G

SEQ ID NO:39    318 asvqakvrtvlaqalvespprveavvvvrhtg-qeilwnegrdrwshdll
SEQ ID NO:130   266 ----------------qsppikrscpdv------qiswnqgidlwwhelm
SEQ ID NO:131   221 ----------------ekatgvesvivlprlglkdvpmtegrdywwnklm
                                   ESPP VE V VV   G   I WNEGRD W H L

SEQ ID NO:39    367 daalakilanaraagfdvhsendllnlpddqliralyasipcep--vdae
SEQ ID NO:130   294 qea-------------------------gde----------cepewcdae
SEQ ID NO:131   255 q------------------------gippn-------ayiepep--vese
                      A                        P D       A I CEP  VDAE

SEQ ID NO:39    415 ypmfiiytsgstgkpkgvihvhggyvagvvhtlrvsfdaepgdtiyviad
SEQ ID NO:130   309 dplfilytsgstgkpkgvvhtvggymlyvattfkyvfdfhaedvfwctad
SEQ ID NO:131   272 hpsfilytsgttgkpkgivhdtggwavhvyatmkwvfdirdddifwctad
                     P FI YTSGSTGKPKGV H  GGY   V  T    FD    D     AD

SEQ ID NO:39    465 pgwitgqsymltatmagrltgviaegsplfpsagryasiierygvqifka
SEQ ID NO:130   359 igwitghsyvtygplangatsvlfegiptypdvnrlwsivdkykvtkfyt
SEQ ID NO:131   322 igwvtghsyvvlgpllmgateviyegapdypqpdrwwsiierygvtifyt
                     GWITG SY      A   T VI EG P  P   R  SIIERYGV IF
```

```
SEQ ID NO:39    515 gvtflktvmsnpqnvedvrlydmhslrvatfcaepvspavqqfgmqimtp
SEQ ID NO:130   409 aptairllmkfgd--epvtkhsraslqvlgtvgepinpeawlwyhrvvga
SEQ ID NO:131   372 sptairmfmryge--ewprkhdlstlriihsvgepinpeawrwayrvlgn
                      T     M       E VR  D  SLRV     EP  P

SEQ ID NO:39    565 q---yi---nsywatehggivwthfygnqdfplrpdahtyplpwvmgdvw
SEQ ID NO:130   457 qrcpiv---dtfwqtetgghmltplpgat--pmkpgsatfp----ffgva
SEQ ID NO:131   420 e---kvafgstwwmtetggivishapglylvpmkpgtngpplpgfevdv-
                    Q           W TE GGIV TH  G    P  P   T PLP    DV

SEQ ID NO:39    609 vaetdesgttryrvadfdekgeivitapypyltrtlwgdvpgfeaylrge
SEQ ID NO:130   498 pailnesg----eelegeaegylvfkqpwpgimrtvy------------
SEQ ID NO:131   466 ---vdengnp----appgvkgylvikkpwpgmlhgiw------------
                     A  DESG      A    KG  VI  P P   RT W

SEQ ID NO:39    659 iplrawkgdaerfvktywrrgpngewgyiqgdfaikypdgsftlhgrpdd
SEQ ID NO:130   531 -------gnherfettyfkkfpg---yyvtgdgcqrdqdgyywitgridd
SEQ ID NO:131   496 -------gdperyiktywsrfpg---mfyagdyaikdkdgyiwvlgrade
                           GD ERF KTYW R P     Y  GD AIK  DG     GR DD

SEQ ID NO:39    709 vinvsghrmgteeiegailrdrqitpdspvgncivvgaphrekgltpvaf
SEQ ID NO:130   571 mlnvsghllstaevesalve------heavaeaavvghphpvkgeclycf
SEQ ID NO:131   536 vikvaghrlgtyelesali------shpavaesavvgvpdaikgevpiaf
                    VINVSGHR GT E E A           V    VVG PH  KG  P AF

SEQ ID NO:39    759 iqpapgrhltgadrrrldelvrtekgavsvpedyie-vsafpetrsgkym
SEQ ID NO:130   615 vtlcdghtfspklteelkkqirekigpiatp-dyiqnapglpktrsgkim
SEQ ID NO:131   580 vvlkqgvapsdelrkelrehvrrtigpiaepaqiff-vtklpktrsgkim
                         G       R  L E VR   G    P DYI  V   P TRSGK M

SEQ ID NO:39    808 rrflrnmml-deplgdtttlrnpevleeiaakiaewkrrqrmaeeqqiie
SEQ ID NO:130   664 rrvlrkiaqndhdlgdmstvadpsvi-----------------------
SEQ ID NO:131   629 rrllkavat-gaplgdvtt-------------------------------
                    RR LR     D PLGD TT   P V

SEQ ID NO:39    857 ryryfrieyhpptasagklavvtvtnppvnalneraldelntivdhlarr
SEQ ID NO:130   690 ------------------------------------------------
SEQ ID NO:131   647 ------------------------------------------------

SEQ ID NO:39    907 qdvaaivftgqgarsfvagadirqlleeihtveeamalpnnahlafrkie
SEQ ID NO:130   690 ----------------------------------------shl------
SEQ ID NO:131   647 -------------------------ledetsveeak-------------
                                             LE    VEEA       HL

SEQ ID NO:39    957 rmnkpciaaingvalggglefamachyrvadvyaefgqpeinlrllpgyg
SEQ ID NO:130   693 ------------------------------------------------
SEQ ID NO:131   658 ------------------------------------------------

SEQ ID NO:39   1007 gtqrlprllykrnngtgllralemilggrsvpadealklglidaiatgdq
SEQ ID NO:130   693 ------------------------------------------------
SEQ ID NO:131   658 -------------------raye-------------------------
                                      RA E

SEQ ID NO:39   1057 dslslacalaraaigadgqliesaavtqafrhrheqldewrkpdprfadd
SEQ ID NO:130   693 ----------------------------fshr----------------
SEQ ID NO:131   662 ------------------------------------------------
                                              F HR
```

```
SEQ ID NO:39   1107 elrsiiahprieriirqahtvgrdaavhraldairygiihgfeaglehea
SEQ ID NO:130   697 ----------------------------------------------------
SEQ ID NO:131   662 ----------------------------------------------------

SEQ ID NO:39   1157 klfaeavvdpnggkrgirefldrqsaplptrrplitpeqeqllrdqkell
SEQ ID NO:130   697 ----------------------------------------------------
SEQ ID NO:131   662 ----------------------------------------------------

SEQ ID NO:39   1207 pvgspffpgvdripkwqyaqavirdpdtgaaahgdpivaekqiivpverp
SEQ ID NO:130   697 ----------------------------------------------------
SEQ ID NO:131   662 ----------------------------------------------------

SEQ ID NO:39   1257 ranqaliyvlasevnfndiwaitgipvsrfdehdrdwhvtgsggigliva
SEQ ID NO:130   697 ----------------------------------------------------
SEQ ID NO:131   662 ----------------------------------------------------

SEQ ID NO:39   1307 lgeearregrlkvgdlvaiysgqsdllsplmgldpmaadfviqgndtpdg
SEQ ID NO:130   697 ----------------------------------------------------
SEQ ID NO:131   662 ----------------------------------------------------

SEQ ID NO:39   1357 shqqfmlaqapqclpiptdmsieaagsyilnlgtiyralfttlqikagrt
SEQ ID NO:130   697 ----------cl---------------------------tiq------
SEQ ID NO:131   662 ----------------------------------------eika---
                                CL                          T QIKA

SEQ ID NO:39   1407 ifiegaatgtgldaarsaarnglrvigmvssssrastllaagahgainrk
SEQ ID NO:130   702 ----------------------------------------------------
SEQ ID NO:131   666 ----------------------------------------------------

SEQ ID NO:39   1457 dpevadcftrvpedpsawaaweaagqpllamfraqndgrladyvvshage
SEQ ID NO:130   702 ----------------------------------------------------
SEQ ID NO:131   666 ----------------------------------------------------

SEQ ID NO:39   1507 tafprsfqllgeprdghiptltfygatsgyhftflgkpgsasptemlrra
SEQ ID NO:130   702 ----------------------------------------------------
SEQ ID NO:131   666 ----------------------------------------------------

SEQ ID NO:39   1557 nlrageavliyygvgsddlvdtggleaieaarqmgarivvvtvsdaqref
SEQ ID NO:130   702 ----------------------------------------------------
SEQ ID NO:131   666 ----------------------------------------------------

SEQ ID NO:39   1607 vlslgfgaalrgvvslaelkrrfgdefewprtmpplpnarqdpqglkeav
SEQ ID NO:130   702 ----------------------------------------------------
SEQ ID NO:131   666 --------------------------emart-------------------
                                           E  RT

SEQ ID NO:39   1657 rrfndlvfkplgsavgvflrsadnprgypdliieraahdalavsamlikp
SEQ ID NO:130   702 ----------------------------------------------------
SEQ ID NO:131   671 ----------------------------------------------------
```

```
SEQ ID NO:39   1707 ftgrivyfediggrrysffapqiwvrqrriymptaqifgthlsnayeilr
SEQ ID NO:130   702 ------------------------------------------------------
SEQ ID NO:131   671 ------------------------------------------------------

SEQ ID NO:39   1757 lndeisaglltitepavvpwdelpeahqamwenrhtaatyvvnhalprlg
SEQ ID NO:130   702 ------------------------------------------------------
SEQ ID NO:131   671 ------------------------------------------------------

SEQ ID NO:39   1807 lknrdelyeawtager
SEQ ID NO:130   702 ----------------
SEQ ID NO:131   671 ----------------
```

**Figure 41**

```
SEQ ID NO:39      1 midtaplappraprsnpirdrvdweaqraaaladpgafhgaiartvihwy
SEQ ID NO:132     1 ------------------------------------------------
SEQ ID NO:133     1 ------------------------------------------------

SEQ ID NO:39     51 dpqhhcwirfnessqrwegldaatgapvtvdypadyqpwqqafddseapf
SEQ ID NO:132     1 ------------------------------------------------
SEQ ID NO:133     1 ------------------md----------------------------
                                      D

SEQ ID NO:39    101 yrwfsggltnacfnevdrhvmmgygdevayyfegdrwdnslnngrggpvv
SEQ ID NO:132     1 ------------------------------------melnn-------
SEQ ID NO:133     3 ----------fnnv-----------------------------------
                            FN V                          LNN

SEQ ID NO:39    151 qetitrrrllvevvkaaqvlrdlglkkgdrialnmpnimpqiyyteaakr
SEQ ID NO:132     6 ------------------------------------------------
SEQ ID NO:133     7 ----------------------llnkddgial----------------
                                        L K D IAL

SEQ ID NO:39    201 lgilytpvfggfsdktlsdrihnagarvvitsdgayrnaqvvpykeaytd
SEQ ID NO:132     6 ------------------------------------------------
SEQ ID NO:133    17 ------------------------------------------------

SEQ ID NO:39    251 qaldkyipvetaqaivaqtlatlpltesqrqtiiteveaalageitvers
SEQ ID NO:132     6 --------------------------------vileke------------
SEQ ID NO:133    17 ------------------------------------------------
                                              I E E

SEQ ID NO:39    301 dvmrgvgsalaklrdldasvqakvrtvlaqalvespprveavvvvrhtgq
SEQ ID NO:132    12 ------------------------------------------------
SEQ ID NO:133    17 ------------------------------------------------

SEQ ID NO:39    351 eilwnegrdrwshdlldaalakilanaraagfdvhsendllnlpddqlir
SEQ ID NO:132    12 ------------------------------------------------
SEQ ID NO:133    17 ----------------------iiin------------------------
                                       I  N

SEQ ID NO:39    401 alyasipcepvdaeypmfiiytsgstgkpkgvihvhggyvagvvhtlrvs
SEQ ID NO:132    12 ------------------------------------------------
SEQ ID NO:133    21 ------------------------------------------------

SEQ ID NO:39    451 fdaepgdtiyviadpgwitgqsymltatmagrltgviaegsplfpsagry
SEQ ID NO:132    12 ------------------------------------------------
SEQ ID NO:133    21 ------------------------------------------------

SEQ ID NO:39    501 asiierygvqifkagvtflktvmsnpqnvedvrlydmhslrvatfcaepv
SEQ ID NO:132    12 ------------------------------------------------
SEQ ID NO:133    21 ------------------------------------------------
```

```
SEQ ID NO:39    551 spavqqfgmqimtpqyinsywatehggivwthfygnqdfplrpdahtypl
SEQ ID NO:132    12 ----------------------------------------------------
SEQ ID NO:133    21 ----------------------------------------rpka-----
                                                            RP A

SEQ ID NO:39    601 pwvmgdvwvaetdesgttryrvadfdekgeivitapypyltrtlwgdvpg
SEQ ID NO:132    12 ----------------------------------------------------
SEQ ID NO:133    25 ----------------------------------------------------

SEQ ID NO:39    651 feaylrgeiplrawkgdaerfvktywrrgpngewgyiqgdfaikypdgsf
SEQ ID NO:132    12 ----------------------------------------------------
SEQ ID NO:133    25 ----------------------------------------------------

SEQ ID NO:39    701 tlhgrpddvinvsghrmgteeiegailrdrqitpdspvgncivvgaphre
SEQ ID NO:132    12 ----------------------------------------------------
SEQ ID NO:133    25 ----------------------------------------------------

SEQ ID NO:39    751 kgltpvafiqpapgrhltgadrrrldelvrtekgavsvpedyievsafpe
SEQ ID NO:132    12 ----------------------------------------------------
SEQ ID NO:133    25 ----------------------------------------------------

SEQ ID NO:39    801 trsgkymrrflrnmmldeplgdtttlrnpevleeiaakiaewkrrqrmae
SEQ ID NO:132    12 ----------------------------------------------------
SEQ ID NO:133    25 ----------------------------------------------------

SEQ ID NO:39    851 eqqiieryryfrieyhpptasagklavvtvtnpp-vnalneraldelnti
SEQ ID NO:132    12 ----------------------gkvavvtinrpkalnalnsdtlkemdyv
SEQ ID NO:133    25 ----------------------------------lnalnyetlkeldsv
                                        GK AVVT   P   NALN   L EL

SEQ ID NO:39    900 vdhlarrqdvaaivftgqgarsfvagadirqlleeihtve-eamalpnna
SEQ ID NO:132    40 igeiendsevlaviltgageksfvagadisem-kemntiegrkfgilgnk
SEQ ID NO:133    40 ldivendkeikvliitgsgektfvagadiaemsn--mtpl-eakkfslyg
                       D       V A   TG G  SFVAGADI     E  T E EA     N

SEQ ID NO:39    949 hlafrkiermnkpciaaingvalggglefamachyrvadvyaefgqpein
SEQ ID NO:132    89 --vfrrlellekpviaavngfalgggceiamscdiriassnarfgqpevg
SEQ ID NO:133    87 qkvfrkiemlskpviaavngfalgggcelsmacdiriasknakfgqpevg
                       FRKIE   KP IAA NG ALGGG E AMAC  R A   A FGQPE

SEQ ID NO:39    999 lrllpgyggtqrlprllykrnngtgllralemilggrsvpadealklgli
SEQ ID NO:132   137 lgitpgfggtqrlsrlv-------gmgmakqliftaqnikadealriglv
SEQ ID NO:133   137 lgiipgfsgtqrlprli-------gtskakeliftgdminsdeaykigli
                    L   PG GGTQRLPRL        G   A E I  G    ADEALK GLI

SEQ ID NO:39   1049 daiatgdqdslslacalaraaigadgqliesaavtqafrhrheqldewrk
SEQ ID NO:132   180 n----------------------------------------------------
SEQ ID NO:133   180 skvv--------------------------------------------------

SEQ ID NO:39   1099 pdprfaddelrsiiahprieriirqahtvgrdaavhraldairygiihgf
SEQ ID NO:132   181 ----------------------------------------------------
SEQ ID NO:133   184 --------elsdli------------------------------------
                            EL   I
```

```
SEQ ID NO:39   1149 eagleheaklfaeavvdpnggkrgirefldrqsaplptrrplitpeqeql
SEQ ID NO:132   181 ------------kvveps----------------------------el
SEQ ID NO:133   190 -----eeakklak-----------------------------------
                        EAK  A  VV P                              L

SEQ ID NO:39   1199 lrdqkellpvgspffpgvdripkwqyaqavirdpdtgaaahgdpivaekq
SEQ ID NO:132   189 mntakei-------------------------------------------
SEQ ID NO:133   198 ------------------------------------------kmmsksq
                      KE                                          Q

SEQ ID NO:39   1249 iivpverpranqaliyvlasevnfndiwaitgipvsrfdehdrdwhvtgs
SEQ ID NO:132   196 ---------ank---------------ivsnapva-------------
SEQ ID NO:133   205 i-------------------------------------------------
                    I        AN                      PV

SEQ ID NO:39   1299 ggiglivalgeearregrlkvgdlvaiysgqsdllsplmgldpmaadfvi
SEQ ID NO:132   207 -------vklskqainrgm-------------------------------
SEQ ID NO:133   206 -----aislakeainkg-----------------------------------
                         V L  EA   G

SEQ ID NO:39   1349 qgndtpdgshqqfmlaqapqclpiptdmsieaagsyilnlgtiyralftt
SEQ ID NO:132   219 -------------------qc-didtalafesea---------fgecfst
SEQ ID NO:133   218 ----------------------metdld-------------------
                                    QC  I TD   E                F T

SEQ ID NO:39   1399 lqikagrtifiegaatgtgldaarsaarnglrvigmvssssrastllaag
SEQ ID NO:132   240 edqkdamtafie-------------------------------------
SEQ ID NO:133   224 ---------------tgntieaekfsl----------------------
                       K   T FIE   TG    A

SEQ ID NO:39   1449 ahgainrkdpevadcftrvpedpsawaaweaagqpllamfraqndgrlad
SEQ ID NO:132   252 ------------------------------------------------
SEQ ID NO:133   236 --------------cft-------------------------------
                             CFT

SEQ ID NO:39   1499 yvvshagetafprsfqllgeprdghiptltfygatsgyhftflgkpgsas
SEQ ID NO:132   252 ------------------------------------------------
SEQ ID NO:133   239 ------------------------------------------------


SEQ ID NO:39   1549 ptemlrranlrageavliyygvgsddlvdtggleaieaarqmgarivvvt
SEQ ID NO:132   252 ------------------------------------------------
SEQ ID NO:133   239 ------------------------------------------------


SEQ ID NO:39   1599 vsdaqrefvlslgfgaalrgvvslaelkrrfgdefewprtmpplpnarqd
SEQ ID NO:132   252 -------------------------krk---------------------
SEQ ID NO:133   239 -tddqke------------gmiafse-kr---------------------
                      D Q E             G     E KR

SEQ ID NO:39   1649 pqglkeavrrfndlvfkplgsavgvflrsadnprgypdliieraahdala
SEQ ID NO:132   255 ------------------------------------------ie--------
SEQ ID NO:133   254 ------------------------------------------------
                                                          IE

SEQ ID NO:39   1699 vsamlikpftgrivyfediggrrysffapqiwvrqrriymptaqifgthl
SEQ ID NO:132   257 ------------------------------------------------
SEQ ID NO:133   254 --------------------------apk---------------fgk--
                                            AP                 FG
```

```
SEQ ID NO:39   1749 snayeilrlndeisaglltitepavvpwdelpeahqamwenrhtaatyvv
SEQ ID NO:132   257 ------------------------------------------------
SEQ ID NO:133   260 ------------------------------------------------


SEQ ID NO:39   1799 nhalprlglknrdelyeawtager
SEQ ID NO:132   257 -------gfknr------------
SEQ ID NO:133   260 ------------------------
                          G KNR
```

**Figure 42**

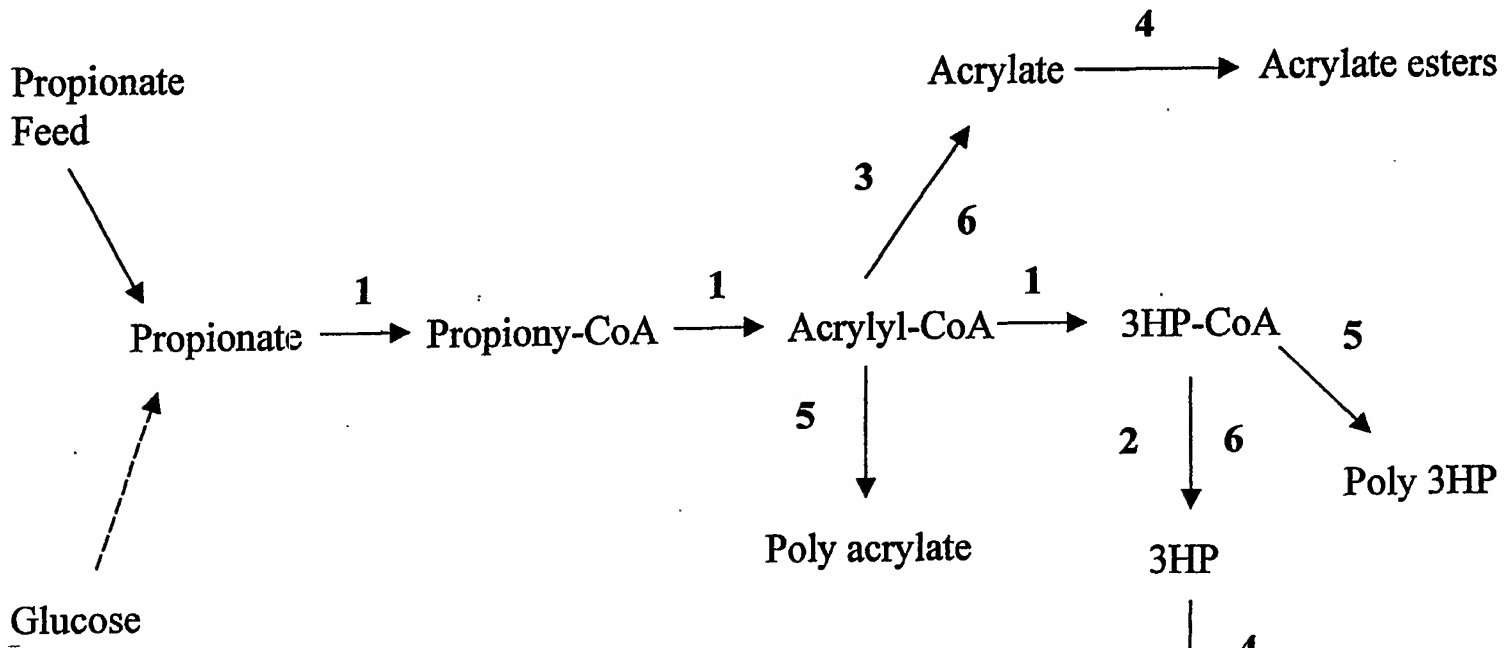
```
SEQ ID NO:39      1 midtaplappraprsnpirdrvdweaqraaaladpgafhgaiartvihwy
SEQ ID NO:134     1 ------------------------------------------------
SEQ ID NO:135     1 ------------------------------------------------

SEQ ID NO:39     51 dpqhhcwirfnessqrwegldaatgapvtvdypadyqpwqqafddseapf
SEQ ID NO:134     1 ----------maasaap----------
SEQ ID NO:135     1 ------------------------------------------------
                                AA  AP

SEQ ID NO:39    101 yrwfsggltnacfnevdrhvmmgygdevayyfegdrwdnslnngrggpvv
SEQ ID NO:134     8 ------------------------------------------------
SEQ ID NO:135     1 ------------------------------------------------

SEQ ID NO:39    151 qetitrrrllvevvkaaqvlrdlglkkgdrialnmpnimpqiyyteaakr
SEQ ID NO:134     8 ------------------------------------------------
SEQ ID NO:135     1 ------------------------------------------------

SEQ ID NO:39    201 lgilytpvfggfsdktlsdrihnagarvvitsdgayrnaqvvpykeaytd
SEQ ID NO:134     8 ----------------------------------------awtg
SEQ ID NO:135     1 ------------------------------------------------
                                                          A T

SEQ ID NO:39    251 qaldkyipvetaqaivaqtlatlpltesqrqtiiteveaalageitvers
SEQ ID NO:134    12 q----------------------------------------taeak
SEQ ID NO:135     1 ----------mtiqtlettalkd----------
                    Q              QTL T  L                  T E

SEQ ID NO:39    301 dvmrgvgsalaklrdldasvqakvrtvlaqalvespprveavvvvrhtgq
SEQ ID NO:134    18 d-----------------------------------------------
SEQ ID NO:135    14 ------------------------------------------------
                    D

SEQ ID NO:39    351 eilwnegrdrwshdlldaalakilanaraagfdvhsendllnlpddqlir
SEQ ID NO:134    19 ------------------------------------------------
SEQ ID NO:135    14 ------------------------------------------------

SEQ ID NO:39    401 alyasipcepvdaeypmfiiytsgstgkpkgvihvhggyvagvvhtlrvs
SEQ ID NO:134    19 ------------------------------------------------
SEQ ID NO:135    14 ------------------------------------------------

SEQ ID NO:39    451 fdaepgdtiyviadpgwitgqsymltatmagrltgviaegsplfpsagry
SEQ ID NO:134    19 ------------------------------------------------
SEQ ID NO:135    14 ------------------------------------------------

SEQ ID NO:39    501 asiierygvqifkagvtflktvmsnpqnvedvrlydmhslrvatfcaepv
SEQ ID NO:134    19 ----------------------------------lyel----------
SEQ ID NO:135    14 ----------------------------------lyei----------
                                                      LY
```

```
SEQ ID NO:39    551 spavqqfgmqimtpqyinsywatehggivwthfygnqdfplrpdahtypl
SEQ ID NO:134    23 ----------------------------------------------------
SEQ ID NO:135    18 ----------------------------------------------------

SEQ ID NO:39    601 pwvmgdvwvaetdesgttryrvadfdekgeivitapypyltrtlwgdvpg
SEQ ID NO:134    23 ----------------------------------------------------
SEQ ID NO:135    18 ----------------------------------------------------

SEQ ID NO:39    651 feaylrgeiplrawkgdaerfvktywrrgpngewgyiqgdfaikypdgsf
SEQ ID NO:134    23 ------geip------------------------------------------
SEQ ID NO:135    18 ------geip------------------------------------------
                        GEIP

SEQ ID NO:39    701 tlhgrpddvinvsghrmgteeiegailrdrqitpdspvgncivvgaphre
SEQ ID NO:134    27 ----------------------------------------------------
SEQ ID NO:135    22 ----------------------------------------------------

SEQ ID NO:39    751 kgltpvafiqpapgrhltgadrrrldelvrtekgavsvpedyievsafpe
SEQ ID NO:134    27 ----------------------------------------------------
SEQ ID NO:135    22 ----------pafhv-------------------------------pk
                             P  H                                 P

SEQ ID NO:39    801 trsgkymrrflrnmmldeplgdtttlrnpevleeiaakiaewkrrqrmae
SEQ ID NO:134    27 ----------------plg-----------hvpakmyawairr----
SEQ ID NO:135    29 t-------------------------------myawsirk----
                    T                 PLG               AK  W  R

SEQ ID NO:39    851 eqqiieryryfrieyhpptasagklavvtvtnppvnalneraldelntiv
SEQ ID NO:134    43 -----erh----------------------------------------
SEQ ID NO:135    38 ----------------------------------------------------
                        ER

SEQ ID NO:39    901 dhlarrqdvaaivftgqgarsfvagadirqlleeihtveeamalpnnahl
SEQ ID NO:134    46 ----------------------------------------------------
SEQ ID NO:135    38 ----------------------------------------------------

SEQ ID NO:39    951 afrkiermnkpciaaingvalggglefamachyrvadvyaefgqpeinlr
SEQ ID NO:134    46 ------------------------------------------gppe----
SEQ ID NO:135    38 -----erhgkp-----------------------------------------
                        ER  KP                                  G PE

SEQ ID NO:39   1001 llpgyggtqrlprllykrnngtgllralemilggrsvpadealklglida
SEQ ID NO:134    50 ----------------------------------------------------
SEQ ID NO:135    44 ----------------------------------------------------

SEQ ID NO:39   1051 iatgdqdslslacalaraaigadgqliesaavtqafrhrheqldewrkpd
SEQ ID NO:134    50 ----------------------------------------------------
SEQ ID NO:135    44 ---------------------------------tqamq-------------
                                                    TQA

SEQ ID NO:39   1101 prfaddelrsiiahprieriirqahtvgrdaavhraldairygiihgfea
SEQ ID NO:134    50 ----------------------qsh-----------------------
SEQ ID NO:135    49 ----------------------------------------------------
                                    Q H
```

```
SEQ ID NO:39   1151 gleheaklfaeavvdpnggkrgirefldrqsaplptrrplitpeqeqllr
SEQ ID NO:134    53 ---------------------------------------------------
SEQ ID NO:135    49 ---------------------------------------------------

SEQ ID NO:39   1201 dqkellpvgspffpgvdripkwqyaqavirdpdtgaaahgdpivaekqii
SEQ ID NO:134    53 -qlevlpv-------------wei---------------gd---------
SEQ ID NO:135    49 ---------------vevvptweige-----------------------
                        Q E LPV       V   P W                 GD

SEQ ID NO:39   1251 vpverpranqaliyvlasevnfndiwaitgipvsrfdehdrdwhvtgsgg
SEQ ID NO:134    65 --------devlvyvmaagvnyngvwaglgepispfdvhkgeyhiagsda
SEQ ID NO:135    60 --------devlvlvmaagvnyngvwaalgepispldghkqpfhiagsda
                              L YV A  VN N  WA  G P S FD H    H  GS

SEQ ID NO:39   1301 iglivalgeearregrlkvgdlvaiysgqsdllsp-lmgldpm-aadfv-
SEQ ID NO:134  107 sgivwkvgakvk---rwkvgdevivhcnqddgddeecnggdpm-fsptqr
SEQ ID NO:135  102 sgivwkvgakvk---rwklgdevvihcnqddgddeecnggdpmfsssqr-
                    G     G       R KVGD V I   Q D       G DPM

SEQ ID NO:39   1348 iqgndtpdgshqqfmlaqapqclpiptdmsieaagsyilnlgtiyralf-
SEQ ID NO:134  153 iwgyetgdgsfaqfcrvqsrqlmarpkhltweeaacytltlatayrmlfg
SEQ ID NO:135  148 iwgyetpdgsfaqfcrvqsrqllprpkhltweesacytltlatayrmlfg
                   I G  TPDGS  QF   Q  Q LP P     E A  Y L L T YR LF

SEQ ID NO:39   1397 -ttlqikagrtifiegaatgtgldaarsaarnglrvigmvssssrastll
SEQ ID NO:134  203 haphtvrpgqnvliwgasgglgvfgvqlcaasganaiavisdeskrdyvm
SEQ ID NO:135  198 hkphelkpgqnvlvwgasgglgvfatqlaavaganaigvvssedkrefvl
                        K G    I GA  G G  A   AA  G   IG VSS S     L

SEQ ID NO:39   1446 aagahgainrkdpevadcftrvpedpsawaaweaagqpllamfraqndgr
SEQ ID NO:134  253 slgakgvinrkd---fdc---w-------------------------------
SEQ ID NO:135  248 smgakavlnrge---fncwgqlpk-------------------------
                     GA G INRKD    DC    P

SEQ ID NO:39   1496 ladyvvshagetafprsfqllgeprdghiptltfygatsgyhftflgkpg
SEQ ID NO:134  269 -------------------------gqlptv-----------------
SEQ ID NO:135  269 ------------------------------------vngpef-------
                                          G  PT        G  F

SEQ ID NO:39   1546 sasptemlrranlrageavliyygvgsddlvdtggleaieaarqmgariv
SEQ ID NO:134  275 ---------------------------------------------------
SEQ ID NO:135  275 ---------------------------------------------------

SEQ ID NO:39   1596 vvtvsdaqrefvlslgfgaalrgvvslaelkrrfgdefewprtmpplpna
SEQ ID NO:134  275 ------------------------------------------------ns
SEQ ID NO:135  275 ----ndymke--------------------srkfgkai-wqit-------
                        D   E                    R FG    W  T     N

SEQ ID NO:39   1646 rqdpqglkeavrrfndlvfkplgsavgvflrsadnprgypdliieraahd
SEQ ID NO:134  277 peyntwlkea-rkfgkaiwditgkgndv----------divfehpgea
SEQ ID NO:135  293 -----gnkdv-----------------------------dmvfehpgeq
                        GLKEA R F        G    V            D   E

SEQ ID NO:39   1696 alavsamlikpftgrivyfediggrrysffapqiwvrqrriymptaqifg
SEQ ID NO:134  314 tfpvstlvakr-ggmivfcagttgfnitfdaryvwmrqkriq-------g
SEQ ID NO:135  308 tfpvsvflvkr-ggmvvicagttgfnltmdarflwmrqkrvq-------g
                     VS  L K   G IV      G    F A   W RQ RI        G
```

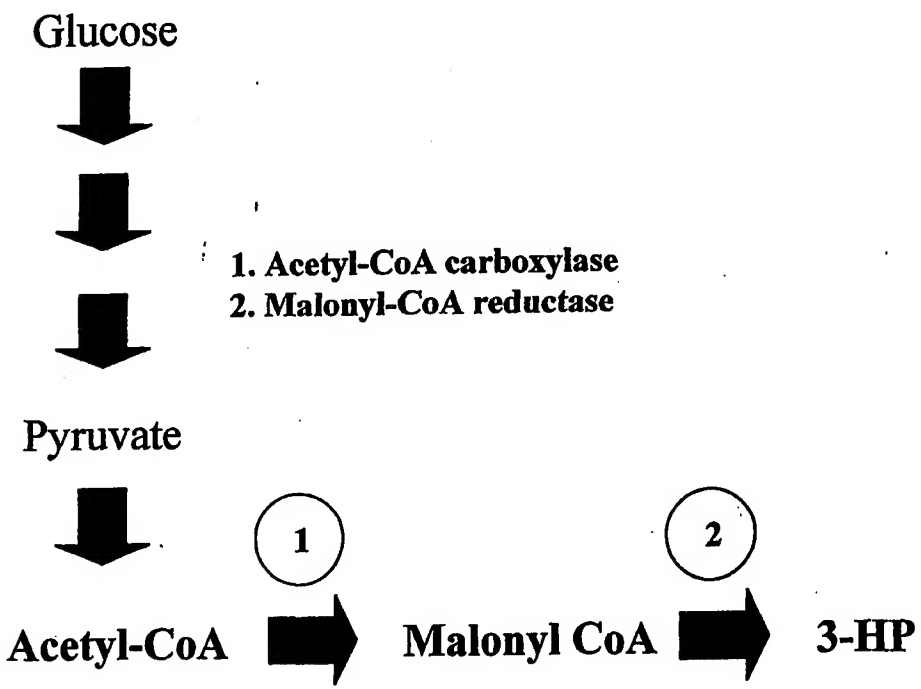
```
SEQ ID NO:39   1746 thlsnayeilrlndeisaglltitepavvpwdelpeahqamwenrhtaat
SEQ ID NO:134   356 shfahlkqasaanqfvmdrrvdpcmsevfpwdkipaahtkmwknqhppgn
SEQ ID NO:135   350 shfanlmqasaanqlvidrrvdpclsevfpwdqipaahekmlanqhlpgn
                     H  N       N              V PWD  P AH  MW N H

SEQ ID NO:39   1796 yvvnhalprlglknrdelyeawtager
SEQ ID NO:134   406 mavlvnstraglrtvedvieagplkam
SEQ ID NO:135   400 mavlvcaqrpglrtfeevqelsgap--
                      V     R GL    E  EA  A
```

# Figure 43




**1. OS17 enzyme**
**2. 3-hydroxyisobutryl or 3-hydroxypropionyl-CoA hydrolyase**
**3. CoA hydrolyase**
**4. Lipase**
**5. Poly hydroxyacid synthase**
**6. CoA transferase**

# Figure 44


Glucose
1. Acetyl-CoA carboxylase
2. Malonyl-CoA reductase
Pyruvate
1
2
Acetyl-CoA
Malonyl CoA
3-HP

# Figure 45

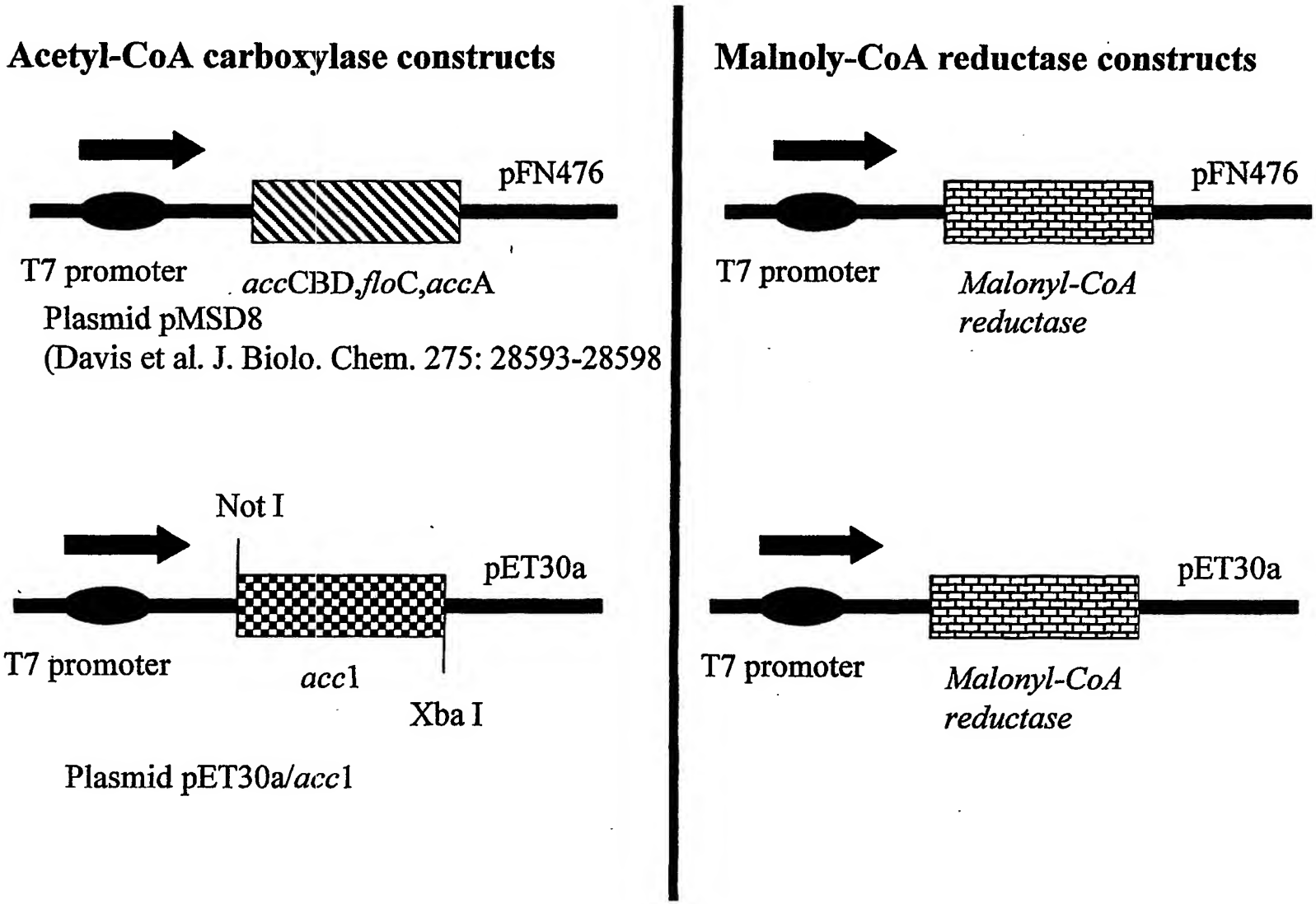

Acetyl-CoA carboxylase constructs
pFN476
T7 promoter
accCBD,floC,accA
Plasmid pMSD8
(Davis et al. J. Biolo. Chem. 275: 28593-28598
Not I
pET30a
T7 promoter
acc1
Xba I
Plasmid pET30a/acc1
Malnoly-CoA reductase constructs
pFN476
T7 promoter
Malonyl-CoA reductase
pET30a
T7 promoter
Malonyl-CoA reductase

Figure 46

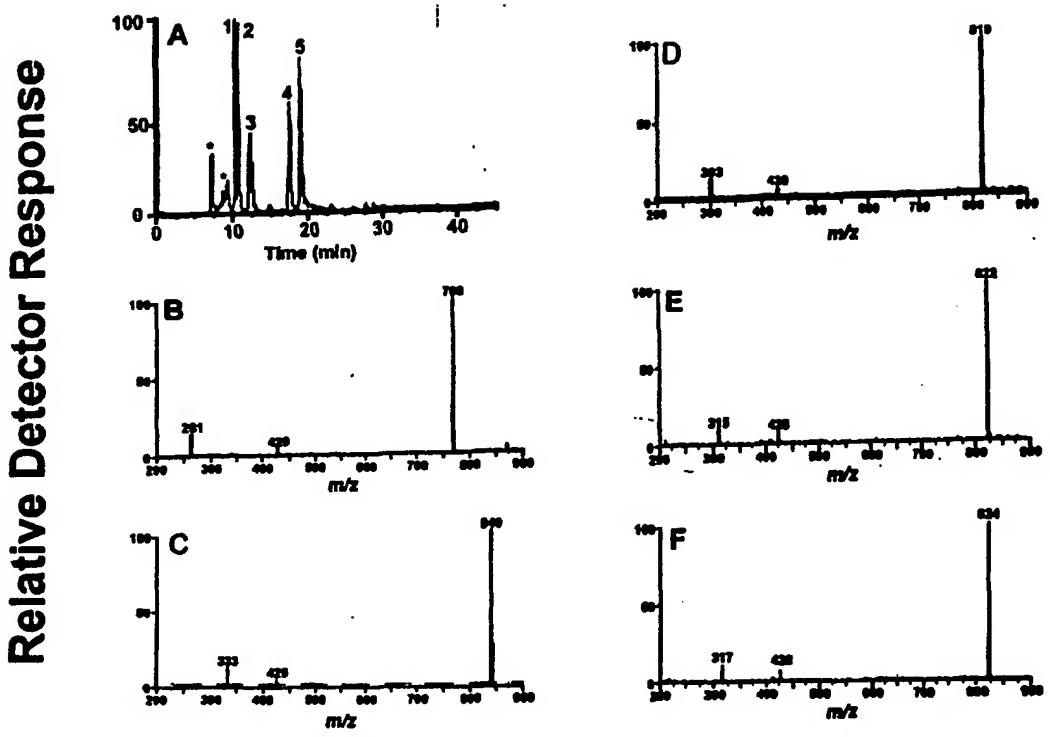


Figure 47

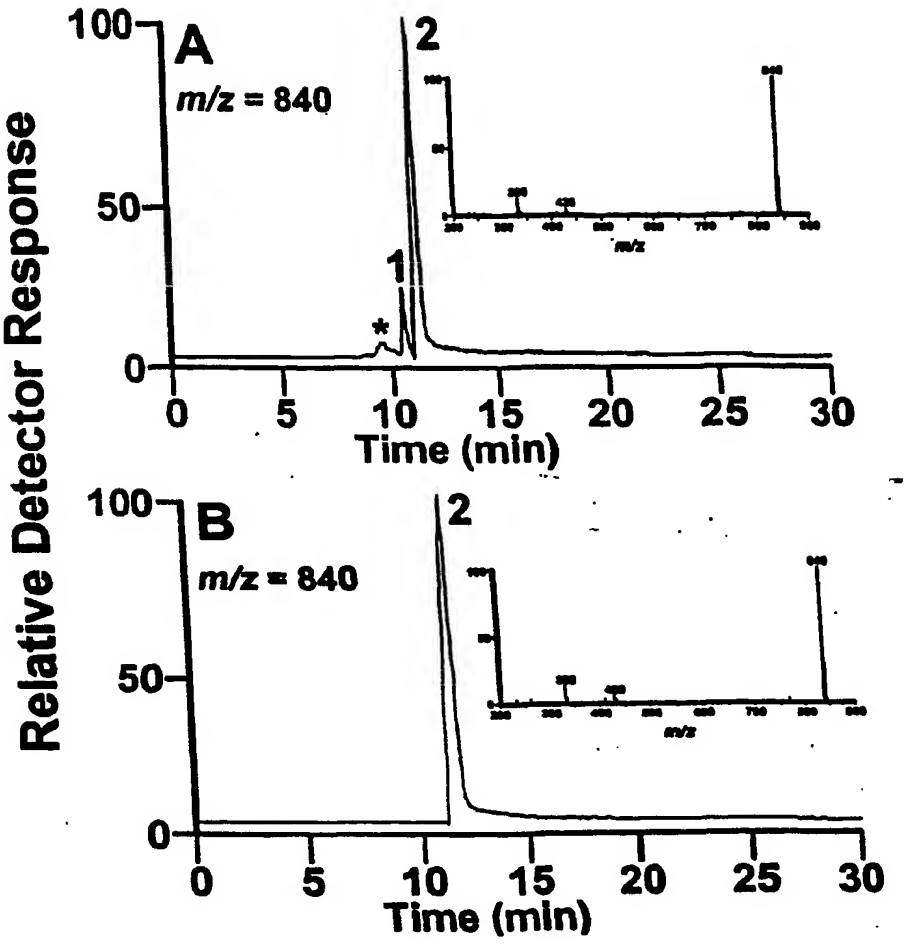

A
m/z = 840
B
m/z = 840
Relative Detector Response
Time (min)
Time (min)
100
50
0
0
5
10
15
20
25
30
1
2
*
m/z

Figure 48

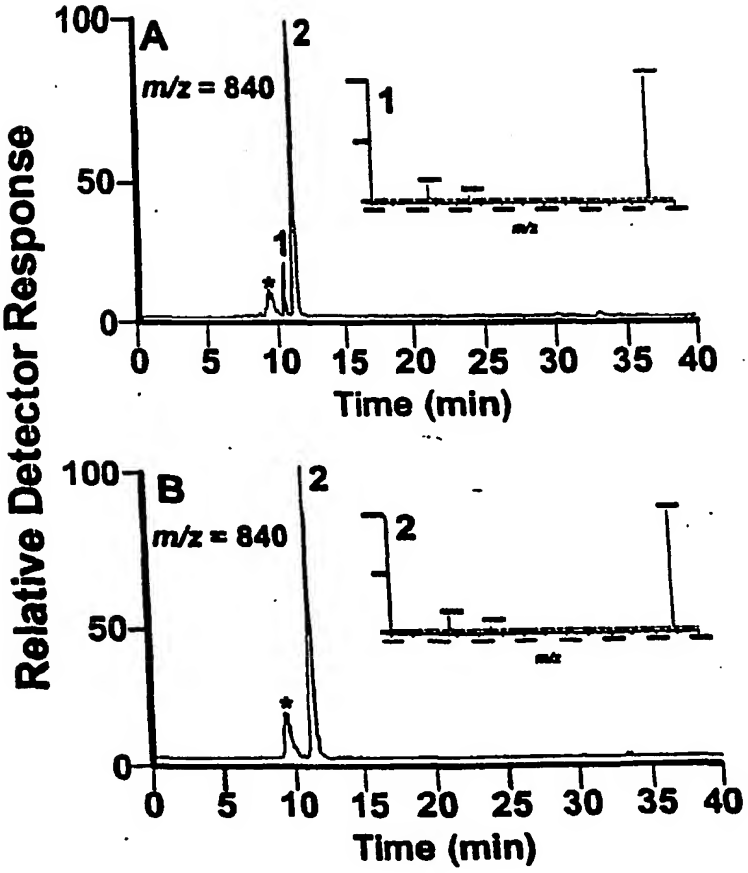

Relative Detector Response
100
50
0
A
m/z = 840
1
2
*
0 5 10 15 20 25 30 35 40
Time (min)
B
m/z = 840
1
2
*
0 5 10 15 20 25 30 35 40
Time (min)

**Figure 49**

ATGGCGACGGGGGAGTCCATGAGCGGAACAGGACGACTGGCAGGAAAGATTGCGTTAATT
ACCGGTGGCGCCGGCAATATCGGCAGTGAATTGACACGTCGCTTTCTCGCAGAGGGAGCG
ACGGTCATTATTAGTGGACGGAATCGGGCGAAGTTGACCGCACTGGCCGAACGGATGCAG
GCAGAGGCAGGAGTGCCGGCAAAGCGCATCGATCTCGAAGTCATGGATGGGAGTGATCCG
GTCGCGGTACGTGCCGGTATCGAAGCGATTGTGGCCCGTCACGGCCAGATCGACATTCTG
GTCAACAATGCAGGAAGTGCCGGTGCCCAGCGTCGTCTGGCCGAGATTCCACTCACTGAA
GCTGAATTAGGCCCTGGCGCCGAAGAGACGCTTCATGCCAGCATCGCCAATTTACTTGGT
ATGGGATGGCATCTGATGCGTATTGCGGCACCTCATATGCCGGTAGGAAGTGCGGTCATC
AATGTCTCGACCATCTTTTCACGGGCTGAGTACTACGGGCGGATTCCGTATGTCACCCCT
AAAGCTGCTCTTAATGCTCTATCTCAACTTGCTGCGCGTGAGTTAGGTGCACGTGGCATC
CGCGTTAATACGATCTTTCCCGGCCCGATTGAAAGTGATCGCATCCGTACAGTGTTCCAG
CGTATGGATCAGCTCAAGGGGCGGCCCGAAGGCGACACAGCGCACCATTTTTTGAACACC
ATGCGATTGTGTCGTGCCAACGACCAGGGCGCGCTTGAACGTCGGTTCCCCTCCGTCGGT
GATGTGGCAGACGCCGCTGTCTTTCTGGCCAGTGCCGAATCCGCCGCTCTCTCCGGTGAG
ACGATTGAGGTTACGCACGGAATGGAGTTGCCGGCCTGCAGTGAGACCAGCCTGCTGGCC
CGTACTGATCTGCGCACGATTGATGCCAGTGGCCGCACGACGCTCATCTGCGCCGGCGAC
CAGATTGAAGAGGTGATGGCGCTCACCGGTATGTTGCGTACCTGTGGGAGTGAAGTGATC
ATCGGCTTCCGTTCGGCTGCGGCGCTGGCCCAGTTCGAGCAGGCAGTCAATGAGAGTCGG
CGGCTGGCCGGCGCAGACTTTACGCCTCCCATTGCCTTGCCACTCGATCCACGCGATCCG
GCAACAATTGACGCTGTCTTCGATTGGGCCGGCGAGAATACCGGCGGGATTCATGCAGCG
GTGATTCTGCCTGCTACCAGTCACGAACCGGCACCGTGCGTGATTGAGGTTGATGATGAG
CGGGTGCTGAATTTTCTGGCCGATGAAATCACCGGGACAATTGTGATTGCCAGTCGCCTG
GCCCGTTACTGGCAGTCGCAACGGCTTACCCCCGGCGCACGTGCGCGTGGGCCGCGTGTC
ATTTTTCTCTCGAACGGTGCCGATCAAAATGGGAATGTTTACGGACGCATTCAAAGTGCC
GCTATCGGTCAGCTCATTCGTGTGTGGCGTCACGAGGCTGAACTTGACTATCAGCGTGCC
AGCGCCGCCGGTGATCATGTGCTGCCGCCGGTATGGGCCAATCAGATTGTGCGCTTCGCT
AACCGCAGCCTTGAAGGGTTAGAATTTGCCTGTGCCTGGACAGCTCAATTGCTCCATAGT
CAACGCCATATCAATGAGATTACCCTCAACATCCCTGCCAACATTAGCGCCACCACCGGC
GCACGCAGTGCATCGGTCGGATGGGCGGAAAGCCTGATCGGGTTGCATTTGGGGAAAGTT
GCCTTGATTACCGGTGGCAGCGCCGGTATTGGTGGGCAGATCGGGCGCCTCCTGGCTTTG
AGTGGCGCGCGCGTGATGCTGGCAGCCCGTGATCGGCATAAGCTCGAACAGATGCAGGCG
ATGATCCAATCTGAGCTGGCTGAGGTGGGGTATACCGATGTCGAAGATCGCGTCCACATT
GCACCGGGCTGCGATGTGAGTAGCGAAGCGCAGCTTGCGGATCTTGTTGAACGTACCCTG
TCAGCTTTTGGCACCGTCGATTATCTGATCAACAACGCCGGGATCGCCGGTGTCGAAGAG
ATGGTTATCGATATGCCAGTTGAGGGATGGCGCCATACCCTCTTCGCCAATCTGATCAGC
AACTACTCGTTGATGCGCAAACTGGCGCCGTTGATGAAAAAACAGGGTAGCGGTTACATC
CTTAACGTCTCATCATACTTTGGCGGTGAAAAAGATGCGGCCATTCCCTACCCCAACCGT
GCCGATTACGCCGTCTCGAAGGCTGGTCAGCGGGCAATGGCCGAAGTCTTTGCGCGCTTC
CTTGGCCCGGAGATACAGATCAATGCCATTGCGCCGGGTCCGGTCGAAGGTGATCGCTTG
CGCGGTACCGGTGAACGTCCCGGCCTCTTTGCCCGTCGGGCGCGGCTGATTTTGGAGAAC
AAGCGGCTGAATGAGCTTCACGCTGCTCTTATCGCGGCTGCGCGCACCGATGAGCGATCT
ATGCACGAACTGGTTGAACTGCTCTTACCCAATGATGTGGCCGCACTAGAGCAGAATCCC
GCAGCACCTACCGCGTTGCGTGAACTGGCACGACGTTTTCGCAGCGAAGGCGATCCGGCG
GCATCATCAAGCAGTGCGCTGCTGAACCGTTCAATTGCCGCTAAATTGCTGGCTCGTTTG
CATAATGGTGGCTATGTGTTGCCTGCCGACATCTTTGCAAACCTGCCAAACCCGCCCGAT
CCCTTCTTCACCCGAGCCCAGATTGATCGCGAGGCTCGCAAGGTTCGTGACGGCATCATG
GGGATGCTCTACCTGCAACGGATGCCGACTGAGTTTGATGTCGCAATGGCCACCGTCTAT
TACCTTGCCGACCGCAATGTCAGTGGTGAGACATTCCACCCATCAGGTGGTTTGCGTTAC

GAACGCACCCCTACCGGTGGCGAACTCTTCGGCTTGCCCTCACCGGAACGGCTGGCGGAG
CTGGTCGGAAGCACGGTCTATCTGATAGGTGAACATCTGACTGAACACCTTAACCTGCTT
GCCCGTGCGTACCTCGAACGTTACGGGGCACGTCAGGTAGTGATGATTGTTGAGACAGAA
ACCGGGGCAGAGACAATGCGTCGCTTGCTCCACGATCACGTCGAGGCTGGTCGGCTGATG
ACTATTGTGGCCGGTGATCAGATCGAAGCCGCTATCGACCAGGCTATCACTCGCTACGGT
CGCCCAGGGCCGGTCGTCTGTACCCCCTTCCGGCCACTGCCGACGGTACCACTGGTCGGG
CGTAAAGACAGTGACTGGAGCACAGTGTTGAGTGAGGCTGAATTTGCCGAGTTGTGCGAA
CACCAGCTCACCCACCATTTCCGGGTAGCGCGCAAGATTGCCCTGAGTGATGGTGCCAGT
CTCGCGCTGGTCACTCCCGAAACTACGGCTACCTCAACTACCGAGCAATTTGCTCTGGCT
AACTTCATCAAAACGACCCTTCACGCTTTTACGGCTACGATTGGTGTCGAGAGCGAAAGA
ACTGCTCAGCGCATTCTGATCAATCAAGTCGATCTGACCCGGCGTGCGCGTGCCGAAGAG
CCGCGTGATCCGCACGAGCGTCAACAAGAACTGGAACGTTTTATCGAGGCAGTCTTGCTG
GTCACTGCACCACTCCCGCCTGAAGCCGATACCCGTTACGCCGGGCGGATTCATCGCGGA
CGGGCGATTACCGTGTAA (SEQ ID NO:140)

**Figure 50**

MATGESMSGTGRLAGKIALITGGAGNIGSELTRRFLAEGATVIISGRNRAKLTALAERMQ
AEAGVPAKRIDLEVMDGSDPVAVRAGIEAIVARHGQIDILVNNAGSAGAQRRLAEIPLTE
AELGPGAEETLHASIANLLGMGWHLMRIAAPHMPVGSAVINVSTIFSRAEYYGRIPYVTP
KAALNALSQLAARELGARGIRVNTIFPGPIESDRIRTVFQRMDQLKGRPEGDTAHHFLNT
MRLCRANDQGALERRFPSVGDVADAAVFLASAESAALSGETIEVTHGMELPACSETSLLA
RTDLRTIDASGRTTLICAGDQIEEVMALTGMLRTCGSEVIIGFRSAAALAQFEQAVNESR
RLAGADFTPPIALPLDPRDPATIDAVFDWAGENTGGIHAAVILPATSHEPAPCVIEVDDE
RVLNFLADEITGTIVIASRLARYWQSQRLTPGARARGPRVIFLSNGADQNGNVYGRIQSA
AIGQLIRVWRHEAELDYQRASAAGDHVLPPVWANQIVRFANRSLEGLEFACAWTAQLLHS
QRHINEITLNIPANISATTGARSASVGWAESLIGLHLGKVALITGGSAGIGGQIGRLLAL
SGARVMLAARDRHKLEQMQAMIQSELAEVGYTDVEDRVHIAPGCDVSSEAQLADLVERTL
SAFGTVDYLINNAGIAGVEEMVIDMPVEGWRHTLFANLISNYSLMRKLAPLMKKQGSGYI
LNVSSYFGGEKDAAIPYPNRADYAVSKAGQRAMAEVFARFLGPEIQINAIAPGPVEGDRL
RGTGERPGLFARRARLILENKRLNELHAALIAAARTDERSMHELVELLLPNDVAALEQNP
AAPTALRELARRFRSEGDPAASSSSALLNRSIAAKLLARLHNGGYVLPADIFANLPNPPD
PFFTRAQIDREARKVRDGIMGMLYLQRMPTEFDVAMATVYYLADRNVSGETFHPSGGLRY
ERTPTGGELFGLPSPERLAELVGSTVYLIGEHLTEHLNLLARAYLERYGARQVVMIVETE
TGAETMRRLLHDHVEAGRLMTIVAGDQIEAAIDQAITRYGRPGPVVCTPFRPLPTVPLVG
RKDSDWSTVLSEAEFAELCEHQLTHHFRVARKIALSDGASLALVTPETTATSTTEQFALA
NFIKTTLHAFTATIGVESERTAQRILINQVDLTRRARAEEPRDPHERQQELERFIEAVLL
VTAPLPPEADTRYAGRIHRGRAITV (SEQ ID NO:141)

**Figure 51**

TCTTTCTGGCCAGTGCCGAATCCGCCGCTCTCTCCGGTGAGACGATTGAGGTTACGCACG
GAATGGAGTTGCCGGCCTGCAGTGAGACCAGCCTGCTGGCCCGTACTGATCTGCGCACGA
TTGATGCCAGTGGCCGCACGACGCTCATCTGCGCCGGCGACCAGATTGAAGAGGTGATGG
CGCTCACCGGTATGTTGCGTACCTGTGGGAGTGAAGTGATCATCGGCTTCCGTTCGGCTG
CGGCGCTGGCCCAGTTCGAGCAGGCAGTCAATGAGAGTCGGCGGCTGGCCGGCGCAGACT
TTACGCCTCCCATTGCCTTGCCACTCGATCCACGCG (SEQ ID NO:142)

Figure 52

```
SEQ ID NO:141      1 matgesmsgtgrlagkialitggagnigseltrrflaegatviisgrnra
SEQ ID NO:143      1 ----------mfankvvlvtggssgigaatveafvkegasvafvgrnqa
SEQ ID NO:144      1 ----------mrlegkvclitgaasgigkattllfaqegatviagdiske
SEQ ID NO:145      1 ------------------------------------------------
SEQ ID NO:146      1 ------------------------------mekf--------------
SEQ ID NO:147      1 ----------mrllhkrtlvtggsdgiglaiaeaflsegadvlivgrdaa

SEQ ID NO:141     51 kltalaermqa--e-agvpakridlevmdgsdpvavragieaivarhgqi
SEQ ID NO:143     40 klkevesrcqq--hganilaikadv-----skdeeakiivqqtvdkfgkl
SEQ ID NO:144     41 nldslvk--ea--e--glp----------------------------gkv
SEQ ID NO:145      1 ------------------------------------------------
SEQ ID NO:146      5 ------------------------------------------------
SEQ ID NO:147     41 kleaarqklaalgq-aga----vetssadlatslgvatvveqvketgrpl

SEQ ID NO:141     98 dilvnnagsagaqrrlaeiplteaelgpgaeetlhasianllgmgwhlmr
SEQ ID NO:143     83 dvlvnnagil----rfasv--leptliqtfdetmntnlrpvv----lits
SEQ ID NO:144     57 d------------------------------------------------
SEQ ID NO:145      1 ------------------------------------------------
SEQ ID NO:146      5 ------------------------------------------------
SEQ ID NO:147     86 dipinnagvadl------vpfesv-----seaqfqhsfalnvaaaffltq

SEQ ID NO:141    148 iaaphm-pvgsavinvstifsr-aeyygrip--yvtpkaalnalsqlaar
SEQ ID NO:143    123 laiphliatkgsivnvssilstivripgims--ysvskaamdhftklaal
SEQ ID NO:144     58 ----------------------------p--yv-----lnv-------
SEQ ID NO:145      1 ------------------------------------------------
SEQ ID NO:146      5 ---php-p-------------------------------------------
SEQ ID NO:147    125 gllphf-gagasiinissyfar-kmipkrpssvyslskgalnsltrslaf

SEQ ID NO:141    194 elgargirvntifpgpiesdrirtvfqrmdqlkgrpegdtahhflntmrl
SEQ ID NO:143    171 elapsgvrvnsvnpgpv--------------------------------
SEQ ID NO:144     64 ------------------tdr-----------------------------
SEQ ID NO:145      1 --------------------------mnpmdrqtegqepqh-------
SEQ ID NO:146      9 ------------------------------------------------
SEQ ID NO:147    173 elgprgirvnaiapgtvdt-------------------------------

SEQ ID NO:141    244 crandqgalerrfpsvgdvadaavflasaesaalsgetievthgmelpac
SEQ ID NO:143    188 ---------------ltdia------------------------------
SEQ ID NO:144     67 ------------------------------------------------
SEQ ID NO:145     16 ------------------------------------------------
SEQ ID NO:146      9 ------------fpr-------------------------------------
SEQ ID NO:147    192 -------amrr---------------------------------------

SEQ ID NO:141    294 setsllartdlrtidasgrttlicagdqieevmaltgmlrtcgseviigf
SEQ ID NO:143    193 ------------------------------------------------
SEQ ID NO:144     67 -------------------------dqikev-----------------
SEQ ID NO:145     16 ------------------------------------------------
SEQ ID NO:146     12 --------------------------qtqem-----------------
SEQ ID NO:147    196 -----------ktvd-----------------------------------

SEQ ID NO:141    344 rsaaalaqfeqavnesrrlagadftppialpldprdpatidavfdwagen
SEQ ID NO:143    193 -------------------agsgfspdll-------------------ed
SEQ ID NO:144     73 ------------------------------------------------
SEQ ID NO:145     16 --------------------qdrqpgieskmnp----------------
SEQ ID NO:146     17 ----------------------------------pgttdrm-------
SEQ ID NO:147    200 ------------------------------------------------
```

```
SEQ ID NO:141   394 tggihaavilpatshepapcvievddervlnfladeitgtiviasrlary
SEQ ID NO:143   205 tg----------ahtp--------------------------------
SEQ ID NO:144    73 ----------------------------------------------
SEQ ID NO:145    29 ---------lp-------------------------------------
SEQ ID NO:146    24 ---------------qplp-----------------------------
SEQ ID NO:147   200 ----------------------------------------------

SEQ ID NO:141   444 wqsqrltpgarargprviflsngadqngnvygriqsaaigqlirvwrhea
SEQ ID NO:143   211 ----------------------------------------------
SEQ ID NO:144    73 ----------------------------------------------
SEQ ID NO:145    31 ------------------lsededyrgs--gklk--------------
SEQ ID NO:146    28 ----------------------------------------------dhg
SEQ ID NO:147   200 ----------------------------------------------

SEQ ID NO:141   494 eldyqrasaagdhvlppvwanqivrfanrsleglefacawtaqllhsqrh
SEQ ID NO:143   211 ----------------------------------------------
SEQ ID NO:144    73 ----------------------------------------------
SEQ ID NO:145    45 ----------------------------------------------
SEQ ID NO:146    31 ensyqgsgrlkd-------------------------------------
SEQ ID NO:147   200 ----------------------------------------------

SEQ ID NO:141   544 ineitlnipanisattgarsasvgwaesliglhlgkvalitggsagiggq
SEQ ID NO:143   211 ------------------------------lgkaa------------
SEQ ID NO:144    73 ----------------------------------------------
SEQ ID NO:145    45 ----------------------------gkvaiitggdsgigra
SEQ ID NO:146    43 ----------------------------------kraiitggdsgigra
SEQ ID NO:147   200 ------nlpa----------------------------------------

SEQ ID NO:141   594 igrllalsgarvmlaardrhk-leqmqamiqselaevgytdvedrvhiap
SEQ ID NO:143   216 ----------------------------qse-----------------
SEQ ID NO:144    73 ----------------------------------------------
SEQ ID NO:145    61 aaiafakegadisilyldehsdaeetrkrieke---------nvrcllip
SEQ ID NO:146    58 vaiayaregadvlisylsehd-----damatkalve----eagrkavlaa
SEQ ID NO:147   204 ----------------------------------------------

SEQ ID NO:141   643 gcdvsseaqladlvertlsafgtvdylinnagiagveemvidmpvegwrh
SEQ ID NO:143   219 --------eiadmi-------------------------------------
SEQ ID NO:144    73 -------------vekvvqkygridvlvnnagitr-dallvrmkeedwda
SEQ ID NO:145   102 g-dvgdenhceqavqqtvdhfgkldilvnnaaeqhpqdsilnisteqlek
SEQ ID NO:146    99 g-diqssdhcrrivetavrelggidilvnnaahqatfkniedisdeewel
SEQ ID NO:147   204 ----eakaelkayvers-------------------------------

SEQ ID NO:141   693 tlfanlisnyslmrklaplmkkqgsgyilnvssyfggekdaaipypnrad
SEQ ID NO:143   225 ----------------------------------------------
SEQ ID NO:144   109 vinvnlkgvfnvtqmvvpymikqrngsivnvssvvg-----iygnpgqtn
SEQ ID NO:145   151 tfrtnifsmfhmtkkalphl--qegcaiinttsitayegdtal-----id
SEQ ID NO:146   148 tfrvnmhamfyltkaavphmkk-gsa-iintasi-----nadvpnpilla
SEQ ID NO:147   217 ----------------------------------------------

SEQ ID NO:141   743 yavskagqramaevfarfl-gpe-iqinaiapgpvegdrlrgtgerpglf
SEQ ID NO:143   225 ----------------------------------------------
SEQ ID NO:144   154 yaaskagvigmtktwakelagrn-irvnavapgfie-------------
SEQ ID NO:145   194 ysstkgaivsftrsmaksl-adkgirvnavapgpi--------------
SEQ ID NO:146   191 yattkgaihnfsaglaqml-aergirvnvvapgpi--------------
SEQ ID NO:147   217 yplgrigr-------------------------------------
```

```
SEQ ID NO:141   791 arrarlilenkrlnelhaaliaaartdersmhelvelllpndvaaleqnp
SEQ ID NO:143   225 ------------------------------------------------------
SEQ ID NO:144   189 ------------------------------------------------------
SEQ ID NO:145   228 ----------------------------------------------wtp
SEQ ID NO:146   225 ----------------------------wtplipstmpedtva-dfgk
SEQ ID NO:147   225 ------------------------------------pddlagm----

SEQ ID NO:141   841 aaptalrelarrfrsegdpaassssallnrsiaakllarlhnggyvlpad
SEQ ID NO:143   225 ------------------------------------------------------
SEQ ID NO:144   189 ------------------------------------------------------
SEQ ID NO:145   231 lipatfpe------------------------------------------------
SEQ ID NO:146   244 qvp---------mkrpgqpvelasa-------------------yvmlad
SEQ ID NO:147   232 ------------------------------------------------------

SEQ ID NO:141   891 ifanlpnppdpfftraqidrearkvrdgimgmlylqrmptefdvamatvy
SEQ ID NO:143   225 ----------------------------------------------vy
SEQ ID NO:144   189 ------------------------------------------------------
SEQ ID NO:145   239 ----------------------ekvkq-----------------------
SEQ ID NO:146   266 pmssy------------------------------------------------
SEQ ID NO:147   232 ----------------------------------------------av

SEQ ID NO:141   941 yladrnvsgetfhpsgglryertptggelfglpsperlaelvgstvylig
SEQ ID NO:143   227 lasdk-------------------------------aksvtgscyi--
SEQ ID NO:144   189 ----------------------------tpmteklpekareta----
SEQ ID NO:145   244 ----------------------------hgldtp--------------
SEQ ID NO:146   271 ------vsgatiavtgg-------------------------------
SEQ ID NO:147   234 yla----sdeaawtsggi-----------------------------

SEQ ID NO:141   991 ehltehlnllaraylerygarqvvmivetetgaetmrrllhdhveagrlm
SEQ ID NO:143   242 ------------------------------------------------------
SEQ ID NO:144   204 ---------lsriplgrfgkpe---------------------evaqvi
SEQ ID NO:145   250 ------------------------------------------------------
SEQ ID NO:146   282 ------------------------------------------------------
SEQ ID NO:147   248 ------------------------------------------------------

SEQ ID NO:141  1041 tivagdqieaaidqaitrygrpgpvvctpfrplptvplvgrkdsdwstvl
SEQ ID NO:143   242 ------------------------------------------------------
SEQ ID NO:144   223 lflasdessyvtgqvi---gidgglvi----------------------
SEQ ID NO:145   250 -----------------mgrpgqpv----------------------
SEQ ID NO:146   282 --------------------------kpfl------------------
SEQ ID NO:147   248 ---------favdggyt-------------------------------

SEQ ID NO:141  1091 seaefaelcehqlthhfrvarkialsdgaslalvtpettatstteqfala
SEQ ID NO:143   242 ----------------------mdnglalq-----------------
SEQ ID NO:144   247 ------------------------------------------------------
SEQ ID NO:145   258 ---------eha---------------gayvllasdes-----------
SEQ ID NO:146   286 ------------------------------------------------------
SEQ ID NO:147   256 ------------------------------------------------------

SEQ ID NO:141  1141 nfikttlhaftatigvesertaqrilinqvdltrraraeeprdpherqqe
SEQ ID NO:143   250 ------------------------------------------------------
SEQ ID NO:144   247 ------------------------------------------------------
SEQ ID NO:145   272 ----------------symtgqtihvn----------------------
SEQ ID NO:146   286 ------------------------------------------------------
SEQ ID NO:147   256 ------------------------------------------------------
```

```
SEQ ID NO:141  1191 lerfieavllvtaplppeadtryagrihrgraitv
SEQ ID NO:143   250 ----------------------------------
SEQ ID NO:144   247 ----------------------------------
SEQ ID NO:145   283 ----------------------------ggrfist
SEQ ID NO:146   286 ----------------------------------
SEQ ID NO:147   256 ----------------------ag----------
```

Figure 53

```
SEQ ID NO:140      1 atggcgacgggggagtccatgagcggaacaggacgactggcaggaaagat
SEQ ID NO:148      1 ------------------atga-----------gacttctgcacaagcg
SEQ ID NO:149      1 ------------------atg--------------ttcgcaaataaagt
SEQ ID NO:150      1 ----------------------------atgaggcttgaagggaaag--
SEQ ID NO:151      1 ------------------------------------atggaaa---
SEQ ID NO:152      1 ------------------------------------------------

SEQ ID NO:140     51 tgcgt-taattaccggtggcgccggcaatatcggcagtgaattgacacgt
SEQ ID NO:148     21 cacgc-tggtgaccggcggctc----------------------------
SEQ ID NO:149     18 ggtac-tagtaacaggtggtagctccggtatcggc---------------
SEQ ID NO:150     20 tgtgtctgatcacagg----ggctgcaagcgggatagggaaa-gccacca
SEQ ID NO:151      8 ------------------------aatttccgca-------------ccct
SEQ ID NO:152      1 ------------------------------------------------

SEQ ID NO:140    100 cgctt--tctcgcagagggagcgacggtcattattagtggacggaatcgg
SEQ ID NO:148     42 ------------------------------------ggacggtatcgg
SEQ ID NO:149     52 ----------------gcagctactgt----------------------
SEQ ID NO:150     65 cgcttcttttcgcacaggaag--------------------------ga
SEQ ID NO:151     22 ccctt--tc-------------------------------------
SEQ ID NO:152      1 ------------------------------------------------

SEQ ID NO:140    148 gcgaagttgaccgcactggccgaacggatgcaggcagaggcaggagtgcc
SEQ ID NO:148     54 cc--------tggcaatcgccggaggcgttcctgagcgagg----------
SEQ ID NO:149     63 ------------------------ggaagcattc---------------
SEQ ID NO:150     88 gctacggtgatcg--ctggc-----gat---------------------
SEQ ID NO:151     29 ------------------------------------------------
SEQ ID NO:152      1 ------gtgaacccaatgg----acaga--caaacagaaggacaag----

SEQ ID NO:140    198 ggcaaagcgcatcgatctcgaagtcatggatgggagtgatccggtcgcgg
SEQ ID NO:148     86 ------gcgc--------cgatgtcct----------------------
SEQ ID NO:149     73 ----------------------gttaaggaagg----------------
SEQ ID NO:150    109 --------------atctcga-----------------------------
SEQ ID NO:151     29 ------------------------------------------------
SEQ ID NO:152     35 ----aaccgcagc--------------------------atcagg------

SEQ ID NO:140    248 tacgtgccggtatcgaagcgattgtggcccgtcacggccagatcgacatt
SEQ ID NO:148     99 ------------------gatcgtcggccgtgacgcc------------
SEQ ID NO:149     84 -------cgcttctgtagccttcgtg-----------------------
SEQ ID NO:150    116 --------------------------------aagaaaatctcgactct
SEQ ID NO:151     29 ----------------------------cccgcca---------------
SEQ ID NO:152     50 --------------------------------acagacagccgggcatt

SEQ ID NO:140    298 ctggtcaacaatgcaggaagtgccggtgcccagcgtcgtctggccgagat
SEQ ID NO:148    118 ------------------------------------------gcc-----
SEQ ID NO:149    103 ---------------ggaagaaaccaagccaag----------------
SEQ ID NO:150    133 cttgtgaaagaggcagaagg-----------------------------
SEQ ID NO:151     36 ---------aacccaggaaatgcc--------------------------
SEQ ID NO:152     67 g-agtcaaaaatgaa------tccgctgcc--------------------

SEQ ID NO:140    348 tccactcactgaagctgaattaggccctggcgccgaagagacgcttcatg
SEQ ID NO:148    121 -----------aagct-----------------cgaagccgcgc-----g
SEQ ID NO:149    121 --------cttaag--gaagtag------------agagccgc----tg
SEQ ID NO:150    153 ------------------------------------------------
SEQ ID NO:151     51 ------------------------------------------------
SEQ ID NO:152     90 ------------------------------------------------
```

```
SEQ ID NO:140   398 ccagcatcgccaatttacttggtatgggatggcatctgatgcgtattgcg
SEQ ID NO:148   138 ccagaagc----------------------------------------tggcg
SEQ ID NO:149   144 ccagcagc----------------------------------------
SEQ ID NO:150   153 ----------------actt----------------------------
SEQ ID NO:151    51 ------------------------------------------------cg
SEQ ID NO:152    90 --------------------------gctgtcagaggacgaggattatc

SEQ ID NO:140   448 gcacctcatatgccggtaggaagtgcggtcatcaatgtctcgaccatctt
SEQ ID NO:148   151 gc----------------------------------tcttggcca----
SEQ ID NO:149   152 ----------------------------------atggagccaacatc--
SEQ ID NO:150   157 ------------ccgg--ggaag----------------------------
SEQ ID NO:151    53 gcac------------------------------------------------
SEQ ID NO:152   113 g---------------aggaa----------------------------

SEQ ID NO:140   498 ttcacgggctgagtactacgggcggattccgtatgtcacccctaaagctg
SEQ ID NO:148   162 --------------------ggc----------------------------
SEQ ID NO:149   166 --------ctggctatcaaag--------cagatgtctcc---aaag---
SEQ ID NO:150   166 ------------------------------------------------
SEQ ID NO:151    57 ------------tac--cgatcggatgc---------------agccg
SEQ ID NO:152   119 --------------------gcgg-----------------aaaactg

SEQ ID NO:140   548 ctcttaatgctctatctcaacttgctgcgcgtgagttaggtgcacgtggc
SEQ ID NO:148   165 ----------------------cggcgc--------ggtggagacgtc
SEQ ID NO:149   194 ------------------------------------------acgagga
SEQ ID NO:150   166 ------------------------------------------------
SEQ ID NO:151    76 c------tgcccgat------------------------cacgggg-
SEQ ID NO:152   130 aaaggaa----------aagttg----------------------

SEQ ID NO:140   598 atccgcgttaatacgatctttcccggcccgattgaaagtgatcgcatccg
SEQ ID NO:148   183 gtccgc-------cgatcttgcc----------------------
SEQ ID NO:149   201 agc-----------------------------gaaaatcatcgta----
SEQ ID NO:150   166 ------gttgatccctacgtt----------ttgaacgtgaccg------
SEQ ID NO:151    92 ---------aaaac-------------------------------tcct
SEQ ID NO:152   143 -------------cgatcattactgg----------------------

SEQ ID NO:140   648 tacagtgttccagcgtatggatcagctcaaggggcggcccgaaggcgaca
SEQ ID NO:148   199 ------------------------------------------------
SEQ ID NO:149   217 ------------------------------------------------
SEQ ID NO:150   194 -acag-------------ggatcagataaag---------gaag------
SEQ ID NO:151   101 accagggttcc--------------------ggacgcctgaag------
SEQ ID NO:152   156 ----------------------------------------aggcgaca

SEQ ID NO:140   698 cagcgcaccatttttttgaacaccatgcgattgtgtcgtgccaacgaccag
SEQ ID NO:148   199 ----------------------------------------------accag
SEQ ID NO:149   217 ------------------------------------caacaa----
SEQ ID NO:150   215 -----------ttgtggaaaa-----------agtcgttcaaa-----ag
SEQ ID NO:151   124 ------------------------------------------gacaag
SEQ ID NO:152   164 ------------------------------------------------

SEQ ID NO:140   748 ggcgcgcttgaacgtcggttcccctccgtcggtgatgtggcagacgccgc
SEQ ID NO:148   204 ----------------------cct----------------------
SEQ ID NO:149   223 ------------------------------------------------ac
SEQ ID NO:150   238 tacg---------gtcgaatc------------gatgt------------
SEQ ID NO:151   130 agagc-----------catcatcaccggcgggga-----cagcggcatc
SEQ ID NO:152   164 ------------------------------------------------
```

```
SEQ ID NO:140   798 tgtctttctggccagtgccgaatccgccgctctctccggtgagacgattg
SEQ ID NO:148   207 ------------cggtgtcgcaaccgtcg-tcgagcaggtgaaa------
SEQ ID NO:149   225 tgtc-----gacaagttc----------------------gggaagcttg
SEQ ID NO:150   255 ---------------------------------tctggtga--------
SEQ ID NO:151   163 gg-----cagggccgtggcga-----------tcgcc-------------
SEQ ID NO:152   164 --------------------------------------------------

SEQ ID NO:140   848 aggttacgcacggaatggagttgccggcctgcagtgagaccagcctgctg
SEQ ID NO:148   238 ----------------------------------gagaccggcc-----
SEQ ID NO:149   248 atgt------------------------------------------------
SEQ ID NO:150   263 --------------------------------------------------
SEQ ID NO:151   184 ----tatgcgcgcgagggag-----------------------------c
SEQ ID NO:152   164 ---------gcggaat----------------agggagagc---------

SEQ ID NO:140   898 gcccgtactgatctgcgcacgattgatgccagtggccgcacgacgctcat
SEQ ID NO:148   248 --------------------------------ggccgctcgacattcct
SEQ ID NO:149   252 --------------------gcttgtt-----------aacaacgc----
SEQ ID NO:150   263 ----------------------------------acaacgc----
SEQ ID NO:151   201 ggacgtccttatcagc-------------------------------tat
SEQ ID NO:152   180 --------------------------------------------------

SEQ ID NO:140   948 ctgcgccggcgaccagattgaagaggtgatggcgctcaccggtatgttgc
SEQ ID NO:148   265 at----------caacaatg-----------------ccggt-------
SEQ ID NO:149   267 --------------------------------------------------
SEQ ID NO:150   270 --------------------------------------------------
SEQ ID NO:151   220 ctgag-------cgagcatgacgacgcgatggccaccaaggct-------
SEQ ID NO:152   180 --------------------------------------------------

SEQ ID NO:140   998 gtacctgtgggagtgaagtgatcatcggcttccgttcggctgcggcgctg
SEQ ID NO:148   280 ------------------------------------gtcgccgacctc
SEQ ID NO:149   267 -------tgggatt-------------ctacggttcg------------
SEQ ID NO:150   270 --------gggaat-----------------------------------
SEQ ID NO:151   256 ----ctggtggag-gaag-------------------------------
SEQ ID NO:152   180 --------------------------------------------------

SEQ ID NO:140  1048 gcccagttcgagcaggcagtcaatgagagtcggcggctggccggcgcaga
SEQ ID NO:148   292 gtgccgttcga------------gagcgtcagcg------aggcgca--
SEQ ID NO:149   284 ----------------------cgagtgt---------tctggagccga
SEQ ID NO:150   276 --------------------------------------------------
SEQ ID NO:151   269 ---caggtcgc-aaggccgt---------------gcttgccgccggcga
SEQ ID NO:152   180 ----------agcag-----------------------------------

SEQ ID NO:140  1098 ctttacgcctcccattgccttgccactcgatccacgcgatccggcaacaa
SEQ ID NO:148   321 ------------------------------------gttccagcactcc
SEQ ID NO:149   302 cttta-----------------------ataca-----------aactt
SEQ ID NO:150   276 ---------------------------------------------aacaa
SEQ ID NO:151   300 c---------------------------atccagtcg-tccg---acca
SEQ ID NO:152   185 -----------ctattgcctt-----------------------------

SEQ ID NO:140  1148 ttgacgctg--tcttcgattgggccggcgagaataccggcgggattcatg
SEQ ID NO:148   334 ttcgcgctc---------aatgtggcgg---------cggcg---------
SEQ ID NO:149   317 ttga--------------------------------------------
SEQ ID NO:150   281 gggatgc-----------------------------------gcttcttg
SEQ ID NO:151   318 ttgccgcaggatcgtcgaaacggccgttcgggaactcggcggcat-----
SEQ ID NO:152   195 --------------------------------------------------
```

```
SEQ ID NO:140  1196 cagcggtgattctgcctgctaccagtcacgaaccggcaccgtgcgtgatt
SEQ ID NO:148   358 ---------ttcttcct-------------------cacc----------
SEQ ID NO:149   321 ------tgaaact-----------------------------------
SEQ ID NO:150   296 -------------------------------------------------
SEQ ID NO:151   363 -------------------------------------------------
SEQ ID NO:152   195 ----------------tgcta-----------------------------

SEQ ID NO:140  1246 gaggttgatgatgagcgggtgctgaattttctggccgatgaaatcaccgg
SEQ ID NO:148   370 ---------caggggctgctgccgcattt---------------------
SEQ ID NO:149   328 ----------atgaac--------------------acgaatttac--g
SEQ ID NO:150   296 -----------tgag---------------------gatgaaa-------
SEQ ID NO:151   363 ------------------------------------------------c
SEQ ID NO:152   200 ------------aagaggggggctga-------------------------

SEQ ID NO:140  1296 gacaattgtgattgccagtcgcctggcccgttactggcagtcgcaacggc
SEQ ID NO:148   390 -------------------------------------------------
SEQ ID NO:149   345 tccagttgtcctcatcactagcctg------------------------
SEQ ID NO:150   307 -------------------------------------------------
SEQ ID NO:151   364 gaca---------------------------------------------
SEQ ID NO:152   213 -------------------------------------------------

SEQ ID NO:140  1346 ttacccccggcgcacgtgcgcgtgggccgcgtgtcatttttctctcgaac
SEQ ID NO:148   390 -------cggcgc-----------------------------------c
SEQ ID NO:149   370 -------------------------------------------------
SEQ ID NO:150   307 -------------------------------------------------
SEQ ID NO:151   368 ---------------------------------ttctcgtcaac
SEQ ID NO:152   213 -------------------------------tatctccattctat---ac

SEQ ID NO:140  1396 ggtgccgatcaaaatgggaatgtttacggacgcattcaaagtgccgctat
SEQ ID NO:148   397 ggtgc-----------------------------------------at
SEQ ID NO:149   370 ---------------------------------------------gctat
SEQ ID NO:150   307 ------gaagaagactgggatg-------------------------
SEQ ID NO:151   379 aatgc-----------------------------------------
SEQ ID NO:152   229 ttagacgagca-------------ttcggacgca---------------

SEQ ID NO:140  1446 cggtcagctcattcgtgtgtggcgtcacgaggctgaacttgactatcagc
SEQ ID NO:148   404 cgatca------------------------------------------
SEQ ID NO:149   375 ccctcatttgatt--------gctacaaaagggag---------------
SEQ ID NO:150   323 cggt-----------------------------------aataaac
SEQ ID NO:151   384 -------------------------------------------------
SEQ ID NO:152   250 ---------------------------gagg--------------aaac

SEQ ID NO:140  1496 gtgccagcgccgccggtgatcatgtgctgccgccggtatgggccaatcag
SEQ ID NO:148   410 -------------------------------------------------
SEQ ID NO:149   402 -------------------------------------------------
SEQ ID NO:150   334 gtg---------------------------------------aatc--
SEQ ID NO:151   384 ------------------------------------agcccatcag
SEQ ID NO:152   258 acgcaaacg---------gatc--------------------gaaaaggag

SEQ ID NO:140  1546 attgtgcgcttcgctaaccgcagccttgaagggttagaatttgcctgtgc
SEQ ID NO:148   410 -------------------------------------------------
SEQ ID NO:149   402 -------------------------------------------------
SEQ ID NO:150   341 -------------------------tgaagggt----------------
SEQ ID NO:151   394 -------------------gcgaccttcaag-------------------
SEQ ID NO:152   280 aatgtccgctgc-----------------------------ctgcttatcc
```

```
SEQ ID NO:140  1596 ctggacagctcaattgctccatagtcaacgccatatcaatgagattaccc
SEQ ID NO:148   410 ------------------------------------------------
SEQ ID NO:149   402 ------------------catagttaacg---tatccagtata------
SEQ ID NO:150   349 --------------------gttttcaacg--------------------
SEQ ID NO:151   406 ------------------------------------------------
SEQ ID NO:152   302 cggga-------------------------------------------

SEQ ID NO:140  1646 tcaacatccctgccaacattagcgccaccaccggcgcacgcagtgcatcg
SEQ ID NO:148   410 tcaacatctcttcctattt--------------cgcccgca--------
SEQ ID NO:149   424 ---------ctgtctacaatag-----------------------------
SEQ ID NO:150   359 ------------------------------------------------
SEQ ID NO:151   406 --aacatc---gaagacatcagcgac----------------------
SEQ ID NO:152   307 ------------------------------------------------

SEQ ID NO:140  1696 gtcggatgggcggaaagcctgatcgggttgcatttggggaaagttgcctt
SEQ ID NO:148   437 ------------------------------------------------
SEQ ID NO:149   437 ------------------------------------------------
SEQ ID NO:150   359 ------------------------------------------------
SEQ ID NO:151   427 ----------------------------------gagga----------
SEQ ID NO:152   307 ----gatg-------------------------ttgggga----------

SEQ ID NO:140  1746 gattaccggtggcagcgccggtattggtgggcagatcgggcgcctcctgg
SEQ ID NO:148   437 ------------------------------------------------
SEQ ID NO:149   437 ------------------------------------------------
SEQ ID NO:150   359 ------------------------------------------------
SEQ ID NO:151   432 ------------------------gtggg---------------------
SEQ ID NO:152   318 ------------------------------------------------

SEQ ID NO:140  1796 ctttgagtggcgcgcgcgtgatgctggcagcccgtgatcggcataagctc
SEQ ID NO:148   437 ------------------------------------------------
SEQ ID NO:149   437 ----------------------------------------taa-----
SEQ ID NO:150   359 ------------------------------------------------
SEQ ID NO:151   437 --------------------agctgacattccg---------------c
SEQ ID NO:152   318 ------------------------------------------------

SEQ ID NO:140  1846 gaacagatgcaggcgatgatccaatctgagctggctgaggtggggtatac
SEQ ID NO:148   437 ------------agatgatcc--------------------------
SEQ ID NO:149   440 ----------------------------------------gaatac
SEQ ID NO:150   359 ------------------------------tgactcagatgg-------
SEQ ID NO:151   451 gtcaacatgcacgccatgttc-------------------------tac
SEQ ID NO:152   318 -------------cga-gaaccattgtgaacaagctg-------------

SEQ ID NO:140  1896 cgatgtcgaagatcgcgtccacattgcaccgggctgcgatgtgagtagcg
SEQ ID NO:148   446 ----------------------------------------------cg
SEQ ID NO:149   446 c-----------------------------------------------
SEQ ID NO:150   371 ------------------------------------------------
SEQ ID NO:151   475 c--tgaccaag-------------gcagcgg-----------------
SEQ ID NO:152   341 ------------------------tgca----------------------

SEQ ID NO:140  1946 aagcgcagcttgcggatcttgttgaacgtaccctgtcagcttttggcacc
SEQ ID NO:148   448 aagcg-----------------------------gccatc------cagc
SEQ ID NO:149   447 ------------------------------------------------
SEQ ID NO:150   371 ------------------------------------------------
SEQ ID NO:151   491 ----------tgccgcacatgaagaa----------------gggcagc
SEQ ID NO:152   345 ----gcaaacagtggacc--------------------attttggtaaa
```

```
SEQ ID NO:140  1996 gtcgattatctga-tcaacaacgccgggatcgccggtgtcgaagagatgg
SEQ ID NO:148   463 gtctactccctgt-ccaagggcgc------------------------
SEQ ID NO:149   447 ------------------------------------------------
SEQ ID NO:150   371 ------------------------------------------------
SEQ ID NO:151   514 g---------cga-tcatcaacaccg----------------------
SEQ ID NO:152   370 ctcgat-atcttagtgaacaacgccg----------------------

SEQ ID NO:140  2045 ttatcgatatgccagttgagggatggcgccataccctcttcgccaatctg
SEQ ID NO:148   486 --------------gttga-------------------------------
SEQ ID NO:149   447 ------------------agggattatgtcatacagt-------------
SEQ ID NO:150   371 ------------------------------------------------
SEQ ID NO:151   530 --------cttcca------------tcaatgccgacgttcccaatccg
SEQ ID NO:152   395 ------------------------------------------------ctg

SEQ ID NO:140  2095 atcagcaactactcgttgatgcgcaaactggcgccgttgatgaaaaaaca
SEQ ID NO:148   491 ----------actcgttga-------------------------------
SEQ ID NO:149   466 ------------------------------------------------
SEQ ID NO:150   371 ----------------------------tggtgccctacatgatcaaaca
SEQ ID NO:151   559 atc-----ctactcgcctatgcg-------------------accacca
SEQ ID NO:152   398 aacagcatc-----------------------------------ccca

SEQ ID NO:140  2145 gggtagcggttacatccttaacgtctcatcatactttggcggtgaaaaag
SEQ ID NO:148   500 ------------------------------------------------
SEQ ID NO:149   466 ------------------------------------------------
SEQ ID NO:150   393 gaggaacggttcgatcgtgaacgtctcctctgtcgttgg--------aat
SEQ ID NO:151   584 agggcgcg-----atc------------cacaattt-------------
SEQ ID NO:152   411 ggacag------cattctcaataTttcaaca-----------------

SEQ ID NO:140  2195 atgcggccattccctaccccaaccgtgccgattacgccgtctcgaaggct
SEQ ID NO:148   500 ----------------------------------ccagatcgct
SEQ ID NO:149   466 ------------------------------------gtgtcaaaggct
SEQ ID NO:150   435 atacgggaat-------cctggtcagacgaattacgcggcgtcgaaggcg
SEQ ID NO:151   603 ---------------------cagcgccg--------gtctcg------
SEQ ID NO:152   436 ------------------------------------------------

SEQ ID NO:140  2245 ggtcagcgggcaatggccgaagtctttgcgcgcttccttggcccg---ga
SEQ ID NO:148   510 ggccttcgag-------------------------ctcggcccgcgcgg
SEQ ID NO:149   478 g-----------------------------------------------
SEQ ID NO:150   478 ggagtcataggaatgacc-aagacgt-----------------------
SEQ ID NO:151   617 ----cgcagatgctggccgaa--------------------cgcg---g-
SEQ ID NO:152   436 gaacagctggaa-------aaaacctttcgc------------------

SEQ ID NO:140  2292 gatacagatcaatgccattgcgccgggtccggtcgaaggtgatcgcttgc
SEQ ID NO:148   534 catccgcgtcaacgccatcgcgcccggcacggtcga--------------
SEQ ID NO:149   479 ------------------------------------------------
SEQ ID NO:150   503 ------------------------------gggcgaaggaactcgct---
SEQ ID NO:151   639 gataagagtgaatgtcgtggccccgggcccgatc----------------
SEQ ID NO:152   460 -acaaatatttttttccat-----------------------------

SEQ ID NO:140  2342 gcggtaccggtgaacgtcccggcctctttgcccgtcgggcgcggctgatt
SEQ ID NO:148   570 ------------------------------------------------
SEQ ID NO:149   479 ------------------------------------------------
SEQ ID NO:150   520 ------------------------------------------------
SEQ ID NO:151   673 ----------------------------------tggacgccgctg---
SEQ ID NO:152   477 ------------------------------------------------
```

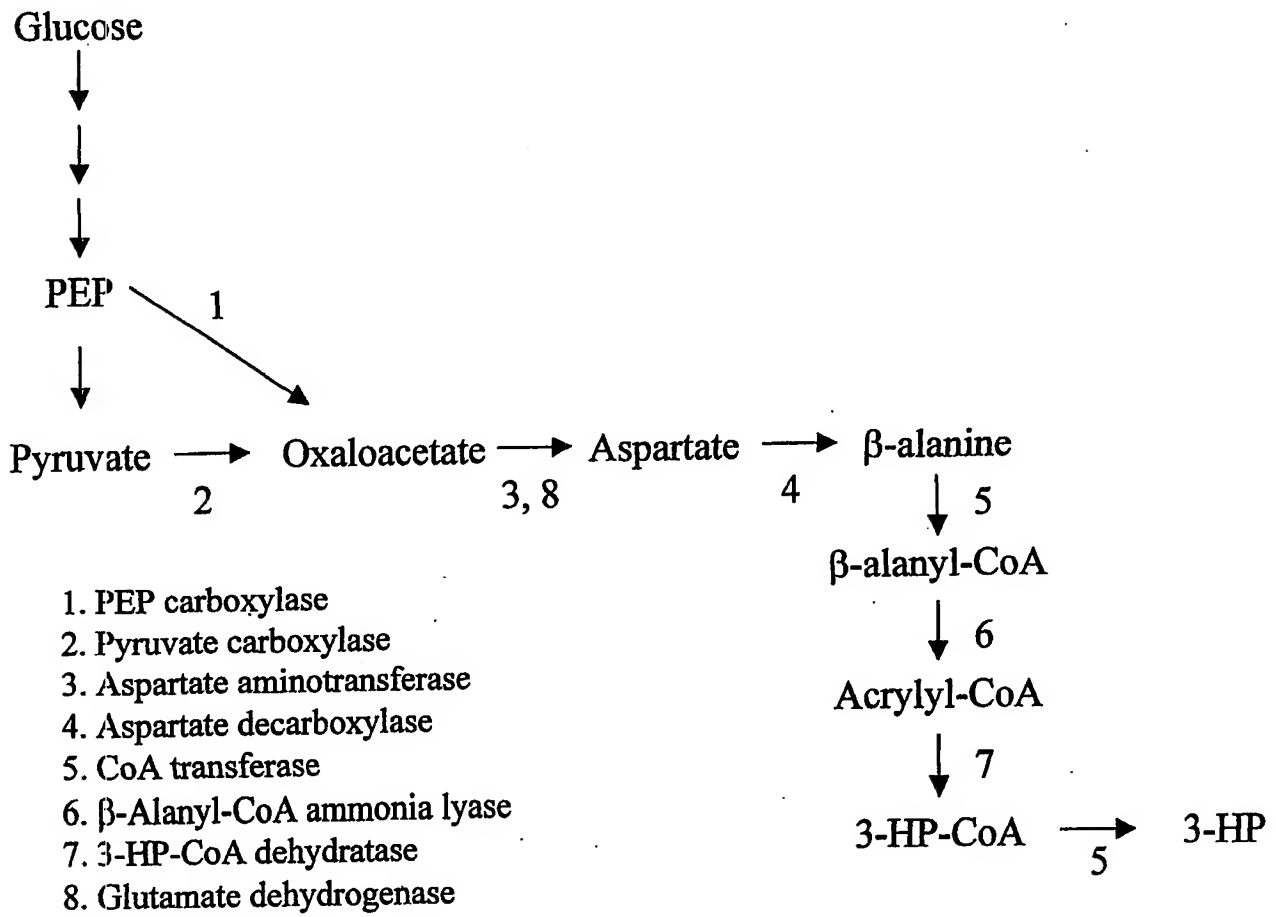
```
SEQ ID NO:140  2392 ttggagaacaagcggctgaatgagcttcacgctgctcttatcgcggctgc
SEQ ID NO:148   570 ------------------------------------------------
SEQ ID NO:149   479 ------------------------------------------------
SEQ ID NO:150   520 ----------------ggaagaaacatcagggtgaac--------gctgt
SEQ ID NO:151   685 ------------------atcccctccaccatgc---------------
SEQ ID NO:152   477 ----------------------gtttca--------------------

SEQ ID NO:140  2442 gcgcaccgatgagcgatctatgcacgaactggttgaactgctcttaccca
SEQ ID NO:148   570 ------------------------------------------------
SEQ ID NO:149   479 ------ctatg---gatcacttcacaaaat--------------------
SEQ ID NO:150   546 g-gcacc----------------------------------------cgga
SEQ ID NO:151   701 -----ccgagga--------------------------------------
SEQ ID NO:152   483 -----------------tatg-acgaa---------------------

SEQ ID NO:140  2492 atgatgtggccgcactagagcagaatcccgcagcacctaccgcgttgcgt
SEQ ID NO:148   570 ----------------------------cacc-------------
SEQ ID NO:149   500 ------tggcagcgttggagctg---------gctccttctggcgtgcga
SEQ ID NO:150   556 ttcat----------------agaaaccccatgac-------------
SEQ ID NO:151   708 ---------------------------------taccg--------
SEQ ID NO:152   492 ------------------------------------gaaagctttgcct

SEQ ID NO:140  2542 gaactggcacgacgttttcgcagcgaaggcgatccggcggcatcatcaag
SEQ ID NO:148   574 ---------------------------gccatgcggcg-------caag
SEQ ID NO:149   535 g------------------------------------------------
SEQ ID NO:150   576 ----------------------cgaaaaacttccag-------aaaaag
SEQ ID NO:151   713 --------tcgccgatttcg-----------------------------
SEQ ID NO:152   505 cacctg-------------------------------------caag

SEQ ID NO:140  2592 cagtgcgctgctgaaccgttcaattgccgctaaattgctggctcgtttgc
SEQ ID NO:148   589 --------------accgt-----------------------------
SEQ ID NO:149   536 ------------tgaac---tcagt-----------------------
SEQ ID NO:150   596 c--------------ccgtgaaacggcc--------------------
SEQ ID NO:151   725 ------------------------------------------------gc
SEQ ID NO:152   515 aggggtg-----------------tgccatta-----------------

SEQ ID NO:140  2642 ataatggtggctatgtgttgcctgccgacatctttgcaaacctgccaaac
SEQ ID NO:148   594 -------------------------cgac---------aacctgcc----
SEQ ID NO:149   546 ------------------------------------caaccctg------
SEQ ID NO:150   610 --------------------------------ctttccaga----------
SEQ ID NO:151   727 aaacaggtgcctatg-----------------------------------
SEQ ID NO:152   530 ttaat------------------acgacat------------------

SEQ ID NO:140  2692 ccgcccgatcccttcttcacccgagcccagattgatcgcgaggctcgcaa
SEQ ID NO:148   606 ------------------------------------------------
SEQ ID NO:149   554 ------gaccagttct---------------------------------
SEQ ID NO:150   619 -------atacc-----------------------------gctgggaa
SEQ ID NO:151   742 ------------------------------------------------aa
SEQ ID NO:152   542 -----cgattaccgctt--------------------------------

SEQ ID NO:140  2742 ggttcgtgacggcatcatggggatgctctacctgcaacggatgccgactg
SEQ ID NO:148   606 ------------------------------------ggccga---
SEQ ID NO:149   564 ---------------------------tac-----------------
SEQ ID NO:150   632 ggtttgggaagccagaagagg-----------------------------
SEQ ID NO:151   744 g------------------------------------------------
SEQ ID NO:152   554 -------------atgaaggggat-----------acgg----------
```

```
SEQ ID NO:140  2792 agtttgatgtcgcaatggccaccgtctattaccttgccgaccgcaatgtc
SEQ ID NO:148   612 ----------------ggcca-----------aggccgaactgaaggcc
SEQ ID NO:149   567 ----tgatatcgc-----------------------------------
SEQ ID NO:150   653 -------tggcgca-----------------------------------
SEQ ID NO:151   745 ----------------------------------cgaccg-------
SEQ ID NO:152   569 ----------------------cgttaattgattattccagcacaaag--

SEQ ID NO:140  2842 agtggtgagaca-ttccacccatcaggtggtttgcgttacgaacgcaccc
SEQ ID NO:148   634 tatg-----------------------------------tcgaacgcagc-
SEQ ID NO:149   576 ------------------------------------------------
SEQ ID NO:150   660 ---ggttatactcttcctcgcatcggacgagtcgagttacg---------
SEQ ID NO:151   751 ------------------------------------------------
SEQ ID NO:152   595 ---ggtgcga-----------------ttgtttcctttacg---------

SEQ ID NO:140  2891 ctaccggtggcgaactcttcggcttgccctcaccggaacggctggcggag
SEQ ID NO:148   649 ---------------tatccgctgggccgcatcgg-ccgtccggacgac
SEQ ID NO:149   576 ----------------------------------------------ag
SEQ ID NO:150   698 ----------------------------tcaccggacagg---------
SEQ ID NO:151   751 --------------------ggccagccc--------------gtggaa
SEQ ID NO:152   616 cgttccatggcgaagtc----gcttgc----------------------

SEQ ID NO:140  2941 ctggtcggaagcacggtctatctgataggtgaacatctgactgaacacct
SEQ ID NO:148   682 ctcgccggcatggcggtttatct--------------------------
SEQ ID NO:149   578 ctggt-----------------------------tctggct--------
SEQ ID NO:150   710 ----------------------tgatag----------------------
SEQ ID NO:151   766 ctcg------cctcggcctatgtcat----------------------
SEQ ID NO:152   639 ------------------------agataaa-----------------

SEQ ID NO:140  2991 taacctgcttgcccgtgcgtacctcgaacgttacggggcacgtcaggtag
SEQ ID NO:148   705 ------------------------------------------------
SEQ ID NO:149   590 ------------------------------------------------
SEQ ID NO:150   716 ------------------------------------------------
SEQ ID NO:151   786 ------------------------------------------------
SEQ ID NO:152   646 ----------------------------------ggca----------

SEQ ID NO:140  3041 tgatgattgttgagacagaaaccggggcagagacaatgcgtcgcttgctc
SEQ ID NO:148   705 ------------------------------------------------
SEQ ID NO:149   590 ----------------------------------------tttctc
SEQ ID NO:150   716 ------------------------------------------------
SEQ ID NO:151   786 ------------------------------------------------
SEQ ID NO:152   650 --------------------------tcagagtgaatgcg----------

SEQ ID NO:140  3091 cacgatcacgtcgaggctggtcggctgatgactattgtggccggtgatca
SEQ ID NO:148   705 ------------------------------------------------
SEQ ID NO:149   596 c---------------------------------------tgatct
SEQ ID NO:150   716 ------------------------------------------------
SEQ ID NO:151   786 ---------------gctgg------------------------------
SEQ ID NO:152   664 ---------------------------------gtggcgcccggt-----

SEQ ID NO:140  3141 gatcgaagccgctatcgaccaggctatcactcgctacggtcgcccagggc
SEQ ID NO:148   705 ---------agccagcgacgaggc-------------------------
SEQ ID NO:149   603 gcttgaag-------------------------------------------
SEQ ID NO:150   716 ------------------------------------------------
SEQ ID NO:151   791 ---cggatccgatgtcga-------------gctac-------------
SEQ ID NO:152   676 --------ccgatttggacaccgct-----------------------
```

```
SEQ ID NO:140  3191 cggtcgtctgtacccccttccggccactgccgacggtaccactggtcggg
SEQ ID NO:148   720 ----------------------------------------------------
SEQ ID NO:149   611 ----------------------------------------------------
SEQ ID NO:150   716 ----------------------------------------------------
SEQ ID NO:151   811 ----------------------------------------------------
SEQ ID NO:152   693 ---------------tattccgg-------cgacattccctgagg-----

SEQ ID NO:140  3241 cgtaaagacagtgactggagcacagtgttgagtgaggctgaatttgccga
SEQ ID NO:148   720 -----------ggcctgga----------------------------cga
SEQ ID NO:149   611 -----atacaggg-------------------------------------
SEQ ID NO:150   716 ---------------------------------------gaat-------
SEQ ID NO:151   811 ----------gtgtcaggcgca----------------------------
SEQ ID NO:152   716 -----aaaaagtga-aacagcac-----------ggcttggataccca

SEQ ID NO:140  3291 gttgtgcgaacaccagctcacccaccatttccgggtagcgcgcaagattg
SEQ ID NO:148   731 gcggtgggatc-------------------------------------tttg
SEQ ID NO:149   619 ---------------gctcatacaccgt----------------------
SEQ ID NO:150   720 ----------------------------------------------------
SEQ ID NO:151   823 ---------------------------------------------acgattg
SEQ ID NO:152   748 ---atgggaagaccgggacagcc----------ggttgagc---------

SEQ ID NO:140  3341 ccctgagtgatggtgc-cagtctcgcgctggtcactcccgaaactacggc
SEQ ID NO:148   746 ccgtg---gatggt-------------------------------------
SEQ ID NO:149   632 ---tggggaaagctgcgcagtct-----------------------------
SEQ ID NO:150   720 -------agatgg-------------------------------------
SEQ ID NO:151   830 ccgtga------------------------------------------
SEQ ID NO:152   776 ------atgcaggcgc-ctatgttctgctggcgtctgacgaa--------

SEQ ID NO:140  3390 tacctcaactaccgagcaatttgctctggctaacttcatcaaaacgaccc
SEQ ID NO:148   757 ----------------------------------------------------
SEQ ID NO:149   652 -------------gaggagattgct-------------------------
SEQ ID NO:150   726 ----------------------------------------------------
SEQ ID NO:151   836 ----------------------------------------------------
SEQ ID NO:152   811 ------------------------tcttccta------------------

SEQ ID NO:140  3440 ttcacgcttttacggctacgattggtgtcgagagcgaaagaactgctcag
SEQ ID NO:148   757 -------------ggcta--------------------------------
SEQ ID NO:149   664 -------------gatatgatt---------------------------
SEQ ID NO:150   726 ----------------------------------------------------
SEQ ID NO:151   836 ----------------------------------------------------
SEQ ID NO:152   819 ----------------tatga--------------------------cag

SEQ ID NO:140  3490 cgcattctgatcaatcaagtcgatctgacccggcgtgcgcgtgccgaaga
SEQ ID NO:148   762 ----------------------------------------------------
SEQ ID NO:149   673 ------------------gtgtatctg---------gctagtgataaagc
SEQ ID NO:150   726 ------------------------------------------------gg
SEQ ID NO:151   836 ----------------------------------------------------
SEQ ID NO:152   827 ggca----gaccattcatgt-----------------------------gaatg

SEQ ID NO:140  3540 gccgcgtgatccgcacgagcgtcaacaagaactggaacgttttatcgagg
SEQ ID NO:148   762 ----------------------------------------------------
SEQ ID NO:149   696 taagagtgtt-------------------acggggtcctgttat-----
SEQ ID NO:150   728 gcctcgtgat-------------------------------------
SEQ ID NO:151   836 ----------------------------------------------------
SEQ ID NO:152   848 gcggc--------------------------------cgttttat-----
```

```
SEQ ID NO:140  3590 cagtcttgctggtcactgcaccactcccgcctgaagccgatacccgttac
SEQ ID NO:148   762 ------------------------------------------------
SEQ ID NO:149   721 -----atcatggacaatg---gactcgcgc--------------------
SEQ ID NO:150   738 ------------------------------ctga----------------
SEQ ID NO:151   836 --------------------------ccggcggcaagcc------------
SEQ ID NO:152   861 ------------------------------------------------

SEQ ID NO:140  3640 gccgggcggattcatcgcggacgggcgattaccgtgtaa
SEQ ID NO:148   762 ---------------cacggccggatga-----------
SEQ ID NO:149   743 ----------tgca--------------------gtaa
SEQ ID NO:150   742 ------------------------------------
SEQ ID NO:151   849 ----------------------------tttcctttga-
SEQ ID NO:152   861 ----------ttcaac------------------gtaa
```

# Figure 54


Glucose
PEP
1
Pyruvate
2
Oxaloacetate
3, 8
Aspartate
4
β-alanine
5
β-alanyl-CoA
6
Acrylyl-CoA
7
3-HP-CoA
5
3-HP
1. PEP carboxylase
2. Pyruvate carboxylase
3. Aspartate aminotransferase
4. Aspartate decarboxylase
5. CoA transferase
6. β-Alanyl-CoA ammonia lyase
7. 3-HP-CoA dehydratase
8. Glutamate dehydrogenase

# Figure 55

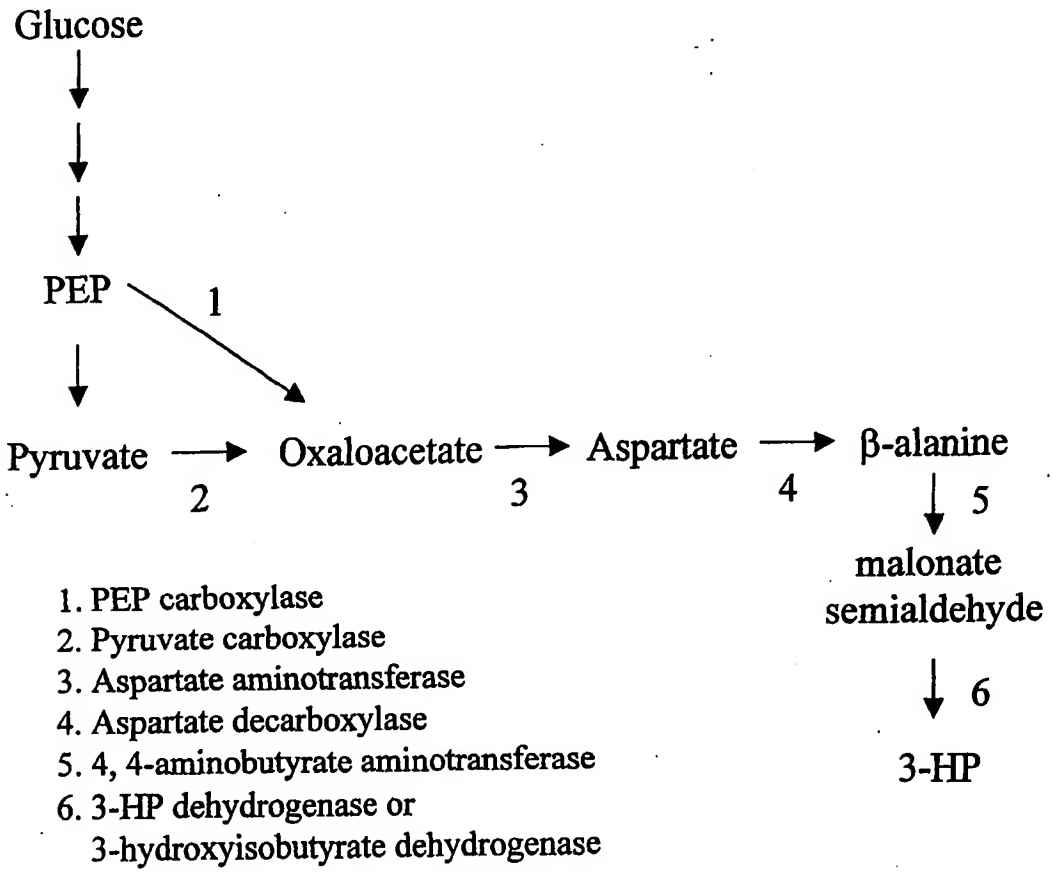

Glucose
PEP
1
Pyruvate
2
Oxaloacetate
3
Aspartate
4
β-alanine
5
malonate
semialdehyde
6
3-HP
1. PEP carboxylase
2. Pyruvate carboxylase
3. Aspartate aminotransferase
4. Aspartate decarboxylase
5. 4, 4-aminobutyrate aminotransferase
6. 3-HP dehydrogenase or
3-hydroxyisobutyrate dehydrogenase

**Figure 56**

```
  1  MVGKKVVHHL MMSAKDAHYT GNLVNGARIV NQWGDVGTEL
 41  MVYVDGDISL FLGYKDIEFT APVYVGDFME YHGWIEKVGN
 81  QSYTCKFEAW KVATMVDITN PQDTRATACE PPVLCGRATG
121  SLFIAKKDQR GPQESSFKER KHPGE  (SEQ ID NO:160)
```

**Figure 57**

```
  1  MVGKKVVHHL MMSAKDAHYT GNLVNGARIV NQWGDVGTEL
 41  MVYVDGDISL FLGYKDIEFT APVYVGDFME YHGWIEKVGN
 81  QSYTCKFEAW KVAKMVDITN PQDTRATACE PPVLCGTATG
121  SLFIAKDNQR GPQESSFKDA KHPQ (SEQ ID NO:161)
```

**Figure 58**

```
  1  ATGGTAGGTA AAAAGGTTGT ACATCATTTA ATGATGAGCG
 41  CAAAAGATGC TCACTATACT GGAAACTTAG TAAACGGCGC
 81  TAGAATTGTG AATCAGTGGG GCGACGTTGG TACAGAATTA
121  ATGGTTTATG TTGATGGTGA CATAAGCTTA TTCTTGGGCT
161  ACAAAGATAT CGAATTCACA GCTCCTGTAT ATGTTGGTGA
201  CTTTATGGAA TACCACGGCT GGATTGAAAA AGTTGGTAAC
241  CAGTCCTATA CATGTAAATT TGAAGCATGG AAAGTTGCAA
281  CAATGGTTGA TATCACAAAT CCTCAGGATA CACGCGCAAC
321  AGCTTGTGAG CCTCCGGTAT TGTGCGGAAG AGCAACGGGT
361  AGTTTGTTCA TCGCAAAAAA AGATCAGAGA GGCCCTCAGG
401  AATCCTCTTT TAAAGAGAGA AAGCACCCCG GTGAATGA
(SEQ ID NO:162)
```

**Figure 59**

```
  1  ATGGTAGGTA AAAAGGTTGT ACATCATTTA ATGATGAGCG
 41  CAAAAGATGC TCACTATACT GGAAACTTAG TAAACGGCGC
 81  TAGAATTGTG AATCAGTGGG GCGACGTAGG TACAGAATTA
121  ATGGTTTATG TTGATGGTGA CATCAGCTTA TTCTTGGGCT
161  ACAAAGATAT CGAATTCACA GCTCCTGTAT ATGTTGGTGA
201  TTTTATGGAA TACCACGGCT GGATTGAAAA AGTTGGCAAC
241  CAGTCCTATA CATGTAAATT TGAAGCATGG AAAGTAGCAA
281  AGATGGTTGA TATCACAAAT CCACAGGATA CACGTGCAAC
321  AGCTTGTGAA CCTCCGGTAC TTTGTGGTAC TGCAACAGGC
361  AGCCTTTTCA TCGCAAAGGA TAATCAGAGA GGTCCTCAGG
401  AATCTTCCTT CAAGGATGCA AAGCACCCTC AATAA
(SEQ ID NO:163)
```